Functional Characterization of the Ewing Sarcoma (EWS) Protein and its Post-Translational Modifications

---

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde (Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Steffen Pahlich

aus

Deutschland

Promotionskomitee

Prof. Dr. Peter Sonderegger (Vorsitz)
Prof. Dr. Heinz Gehring (Leitung der Dissertation)

Zürich, 2007

Die vorliegende Arbeit wurde von der Mathematisch-naturwissenschaftlichen Fakultät der Universität Zürich im Sommersemester 2007 als Dissertation angenommen.

Promotionskomitee:

Prof. Dr. Peter Sonderegger (Vorsitz)

Prof. Dr. Heinz Gehring (Leitung der Dissertation)

The results of this work have been or will be published elsewhere:

List of Publications

Pahlich S, Bschir K, Chiavi C, Belyanskaya L, Gehring H. *Different methylation characteristics of protein arginine methyltransferase 1 and 3 toward the Ewing Sarcoma protein and a peptide*. Proteins. 2005 Oct 1;61(1):164-75.

Pahlich S, Zakaryan RP, Gehring H. *Review: Protein arginine methylation: Cellular functions and methods of analysis*. Biochim Biophys Acta – Proteins and Proteomics. 2006 Dec;1764(12):1890-903

Pahlich S, Zakaryan RP, Gehring H. *Identification of proteins interacting with Protein Arginine Methyltransferase 8: The Ewing Sarcoma (EWS) Protein binds independent of its methylation state.* Submitted for publication.

Pahlich S, Quero L, Zakaryan RP, Gehring H. *Analysis of Ewing Sarcoma (EWS)-binding Proteins: Interaction with Protein-RNA complexes via hnRNP M, U, and the RNA-helicases p68 and 72.* In preparation for publication.

Poster Presentations

S. Pahlich and H. Gehring. *Analysis of Protein Arginine Methylation by Mass Spectrometry: Kinetics and Sequence Preferences of PRMT1 and PRMT3.*
Human Proteome Organization (HUPO) 4th Annual World Congress, August 28th - September 1st, Munich, Germany (2005). Abstract Number: TP087

S. Pahlich and H. Gehring. *Analysis of Protein Arginine Methylation by Mass Spectrometry: Kinetics and Sequence Preferences of PRMT1, PRMT3, and PRMT8*.
Swiss Proteomics Society (SPS) - "Expanding Proteomics", 5-7 December, Zurich, Switzerland (2005). Abstract Number SPS05-3406

Oral Presentations

Lunchtime Seminars on Post-Translational Modifications (Uni/ETH Zurich, WS 06/07), *"The role of protein arginine methylation and phosphorylation in the EWS protein."*

Functional Genomics Center Zurich (FGCZ) User Day 2006; FGCZ User Colloquium, *"Functional characterization of the human EWS protein and its interaction partners."*

Abbreviations

ADMA	asymmetric ω-N^G, N^G-dimethylarginine
AdoHcy	S-adenosyl-L-homocysteine
AdoMet	S-adenosyl-L-methionine
AdOx	adenosine dialdehyde
CARM1	Coactivator-associated arginine methyltransferase 1 (PRMT4)
DHR	degenerated hexa-repeats
EAD	EWS transcriptional activation domain
ESFT	Ewing sarcoma family of tumors
ETS	erythroblastostosis virus-transformed sequence
EWS	Ewing Sarcoma protein
GAR	Glycine-arginine rich regions
hnRNP	heterogeneous nuclear ribonucleoproteins
MALDI	matrix-assisted laser desorption/ionization
MMA	N^G-monomethylarginine
MS	mass spectrometry
NLS	nuclear localization sequence
PRMT	protein arginine methyltransferase
PTM	post-translational modifications
RBD	RNA-binding domain
RRM	RNA-recognition motif
SDMA	symmetric ω-N^G,N'^G-dimethylarginine
TAFII68	TATA-binding protein-associated factor 2N
TLS	Translocated in liposarcoma protein
TOF	time of flight
ZF	zinc finger

Table of Content

Summary

The human Ewing Sarcoma (EWS) protein is a multifunctional RNA-binding protein, which plays a role in transcriptional activation, DNA pairing and repair, RNA splicing, and signal transduction. Beside its mainly nuclear localization, the EWS protein was also found to be associated with the cell membrane, but the mechanism mediating this association is unknown. In the family of Ewing sarcoma tumors, chromosomal translocation of the *EWS* gene results in the expression of fusion proteins of the EWS protein with one of several ETS (erythroblastostosis virus-transformed sequence) transcription factors. The transforming potential of such fusion proteins is repressed in the wild-type EWS protein by the presence of its C-terminal RNA-binding domain (RBD). In previous studies in our lab, the RBD was found to undergo extensively post-translational dimethylation at arginine residues within its arginine-glycin rich regions (RGG boxes 1-3). The contribution of the EWS protein in the above mentioned cellular processes as well as the function of the methylations are investigated within this study.

The first goal of this thesis was to identify the human protein arginine methyltransferases (PRMT) that catalyze the methylation of the EWS protein. Out of nine currently known PRMTs, the three PRMTs 1, 3, and 8, which catalyze the methylation of arginine residues within RGG motifs, were chosen to test for methyltransferase activity towards the EWS protein. PRMT1, the main methyltransferase in human cells, methylated the EWS protein almost completely in vitro (27 of total 30 methylation sites), as determined by mass spectrometry. The cytoplasmic PRMT3 and the membrane-associated PRMT8 methylated only nine and five arginine residues, respectively, mainly within the C-terminal RGG box 3 of the EWS protein. Although all 30 methylation sites are situated in highly similar RGG motifs (RGG boxes 1-3), structural determinants seem to hinder the accessibility of the RGG boxes 1 and 2 for PRMT3 and 8, whereas the C-terminal RGG box 3 of the EWS protein is most accessible for PRMT3 and 8. Additionally, a method to determine so far unknown kinetic parameters of the PRMTs was developed by measuring the methylation of a synthetic peptide with mass spectrometry. The catalytic efficiency of PRMT3 was found to be 12fold higher than that of PRMT1, whereas that of PRMT8 was 30fold lower, supporting the notion that structural rather than kinetic features of the substrate are responsible for the difference in EWS protein methylation by the PRMTs.

The second part of this study shows that PRMT8, the only membrane-associated methyltransferase, interacts with several nuclear, RGG-containing proteins, like the RNA-binding proteins EWS, TLS/FUS, and $TAF_{II}68$, hnRNPs, and RNA-helicases. PRMT8 binds to the RGG box 3 of the EWS protein independent of its methylation state implying an additional role of PRMT8, beside its function as methyltransferase, as an adaptor protein for methylated nuclear proteins like the EWS protein at the membrane.

In the last part of this dissertation, the cellular role of the EWS protein, in particular that of the RBD and its multiple arginine methylations, was investigated by analyzing protein-protein interactions of unmethylated and methylated recombinant EWS protein. In these studies, co-expression of the EWS protein or the RBD and PRMT1 in *E. coli* was successfully established resulting in complete methylation of the EWS protein or the RBD in bacteria. Protein complexes consisting of more than 30 different proteins, mainly heterogeneous nuclear ribonucleoproteins (hnRNPs) and RNA helicases, were found to co-purify with the unmethylated RBD of the EWS protein. The interaction with these RNase-sensitive complexes is mediated either by RNA-binding of the EWS protein or through direct interaction of the EWS protein with hnRNP M, U, and the RNA-helicases p68 and p72. Co-precipitation and co-localization studies of the EWS protein and p68 did not only confirm their interaction, but showed also a re-localization of the EWS protein upon p68 expression from the nucleoplasm to the nucleolar periphery. Such re-localizations were observed for many nuclear proteins upon transcriptional inhibition. Remarkably, all these interactions did not depend on the methylation of the EWS protein although they take place in the extensively methylated domain of the EWS protein.

Methylation of arginine residues has been shown to regulate various cellular functions of proteins like their subcellular localization, functional activation/repression, and protein-protein or protein-nucleic acid interactions. The present study shows that methylation of the EWS protein takes place shortly after translation or even co-translationally, as no unmethylated EWS protein was found, be it in the nucleus, in the cytoplasm, or at the cell-surface. A role of the methylation in the subcellular localization of the EWS protein can be excluded as the EWS methylation is not needed for the nuclear import. As described above, arginine methylations are also dispensable for EWS protein-protein interactions. Thus, the methylation of the EWS protein has to play a role in other functions, such as RNA binding, the fine-tuning of the activation/repression activity of the EWS protein, stabilization of the EWS protein against proteolysis, or it can serve as source for asymmetrically dimethylated

arginine, which is a regulator of nitric oxide synthase (NOS) in the cell. No demethylases are known so far, and the extensive arginine methylation of the EWS might therefore have a static role rather than being a dynamic regulatory element of its function.

Zusammenfassung

Das menschliche Ewing-Sarcoma (EWS)-Protein ist ein multifunktionales RNA-bindendes Protein, welches in diversen zellulären Prozessen wie der transkriptionellen Aktivierung, der Reparatur und Zusammenlagerung von DNA-Strängen, dem Splicing und der Signaltrasduktion eine Rolle spielt. Das EWS-Protein findet sich hauptsächlich im Zellkern, jedoch in geringeren Fraktionen auch an der Zellmembran. Es ist nach wie vor unklar, wie das EWS-Protein an der Zellmembran assoziiert ist. In Ewing-Sarkom-Tumoren führt eine chromosomale Translokation des *EWS*-Gens zu der Expression von Fusionsproteinen, bestehend aus der N-terminalen Domäne des EWS-Proteins (EAD) und der C-terminalen Domäne eines (von mehreren) ETS-Transkriptionsfaktoren. Die Expression solcher Fusionsproteine führt zur Transformation der Zelle und damit zur Tumor-Entstehung. In einer gesunden Zelle wird das Transformationspotential der EAD im Wildtyp-EWS-Protein durch die C-terminale RNA-bindende Domäne (RBD) unterdrückt. Vorhergehende Studien in unserem Labor haben gezeigt, das mehrere Argininreste in dieser RBD post-translational methyliert werden. Dies betrifft jedoch ausschliesslich Argininreste in Arginin-Glycin-reichen Regionen (RGG-Boxen). Die Funktion des EWS-Proteins in den oben genannten zellulären Prozessen wie auch der Beitrag der Methylierungen sind weitgehend unklar und deren Untersuchung ist Gegenstand dieser Studie.

Das erste Ziel dieser Arbeit war die Identifizierung von Protein-Arginin-Methyltransferasen (PRMT), welche die Methylierungen des EWS-Proteins katalysieren. Von bisher neun bekannten menschlichen PRMTs wurden die drei PRMTs 1, 3 und 8 für nähere Untersuchungen ausgewählt, da diese speziell die Methylierung von Argininresten in RGG-Boxen katalysieren. Mittels Massenspektrometrie (MS) konnten wir zeigen, dass PRMT1, die vorherrschende Methyltransferase in der Zelle, das EWS-Protein fast vollständig methyliert (27 von total 30 methylierten Argininresten). Sowohl die cytosolische PRMT3 als auch die zellwand-assoziierte PRMT8 waren nicht in der Lage, das EWS Protein vollständig zu methylieren (nur neun bzw. fünf methylierte Arginine wurden detektiert). Obwohl die Aminosäuresequenzen sämtlicher 30 Methylierungsstellen des EWS Proteins sehr redundant sind, methylieren PRMT3 und 8 hauptsächlich Argininreste innerhalb der C-terminalen RGG-Box 3. Offensichtlich erschweren strukturelle Domänen zwischen den RGG Boxen 1 und 2 den Zugang für PRMT3 und 8. Zusätzlich wurde eine Methode zur Bestimmung bisher unbekannter kinetischer Parameter der PRMTs entwickelt, die ausschliesslich auf MS Daten eines in vitro methylierten Peptids beruht. Die katalytische Effizienz von PRMT3 ist demnach

um ein 12-Faches höher als die von PRMT1. PRMT8 dagegen hat eine 30-fach geringere katalytische Effizienz als PRMT1. Diese Daten unterstützen die Annahme, dass eher die Struktur des Substrats als kinetische Eigenschaften der Enzyme für die unterschiedlichen EWS-Methylierungen verantwortlich ist.

Im zweiten Teil dieser Studie wird gezeigt, dass die zellwand-assoziierte PRMT8 mit diversen Kernproteinen, wie z.B. den RNA-bindenden Proteinen EWS, TLS/FUS, und $TAF_{II}68$, mit heterogenen nuclearen Ribonukleoproteinen (hnRNPs) und RNA-Helikasen interagiert. Die Wechselwirkung zwischen PRMT8 und der C-terminalen RGG-Box 3 des EWS-Proteins findet unabhängig vom Methylierungsgrad des EWS-Proteins statt. Es lässt sich daher vermuten, dass PRMT8 neben seiner Funktion als Methyltransferase eine zusätzliche Rolle als Adapterprotein für methylierte Kernproteine an der Zellmembran einnimmt.

Im letzten Teil dieser Dissertation wurde die zelluläre Funktion des EWS-Proteins, besonders seiner RBD einschliesslich der methylierten Argininreste, untersucht, indem interagierende Proteine des unmethylierten und des methylierten rekombinanten EWS-Proteins bestimmt wurden. Dafür wurde erfolgreich eine Co-Expression des EWS-Proteins oder der RBD und PRMT1 in Bakterien etabliert, womit bereits methyliertes rekombinantes EWS-Protein bzw. RBD in Bakterien hergestellt werden kann. Mehr als 30 verschiedene Proteine, darunter hauptsächlich hnRNPs und RNA-Helikasen, konnten als Interaktionspartner der unmethylierten RBD des EWS-Proteins identifiziert werden. Die Wechselwirkungen des EWS-Proteins mit diesen RNase-empfindlichen Protein-Komplexen wird entweder über eine Wechselwirkung des EWS-Proteins mit RNA oder über eine direkte Bindung zu den Proteinen hnRNP M und U, sowie den RNA-Helikasen p68 und p72, die als direkte Interaktionspartner identifiziert worden sind, erhalten. Die Wechselwirkung zwischen p68 und p72 mit dem EWS-Protein wurde durch Co-Präzipitationen sowie durch Co-Lokalisierung in HEK-Zellen bestätigt. Dabei konnte gezeigt werden, dass sich das EWS Protein bei einer Co-Expression mit p68 vom Kern-Plasma zur Peripherie der Nucleoli verschiebt. Bei gehemmter Transkripition wurde eine solche Umlagerung bereits bei mehreren anderen Kernproteinen beobachtet. Bemerkenswert ist, dass eine Methylierung des EWS-Proteins keinerlei Einfluss auf die Protein-Wechselwirkungen hat, obwohl sämtliche hier gezeigten Wechselwirkungen mit der RBD des EWS-Proteins stattfinden, die sämtliche Methylierungsstellen enthält.

Die Arginin-Methylierung von Proteinen kann mannigfaltige Prozesse wie z.B. die intrazelluläre Lokalisierung von Proteinen, die funktionelle Aktivierung oder Repression von Enzymen, oder die Wechselwirkungen zwischen Proteinen oder Proteinen und Nucleinsäuren steuern. Diese Studie zeigt, dass das EWS-Protein bereits kurz nach oder sogar während der Translation methyliert wird, da in der Zelle bisher kein unmethyliertes EWS-Protein nachgewiesen werden konnte, weder im Zellkern, noch im Cytosol oder an der Zellmembran. Eine Funktion des Methylierungsgrades in der subzellulären Lokalisierung des EWS-Proteins kann ausgeschlossen werden, da sowohl methyliertes als auch unmethyliertes EWS-Protein in den Zellkern transportiert wird. Wie oben beschrieben, sind die Arginin-Methylierungen des EWS-Proteins ebenfalls für Protein-Protein-Wechselwirkungen verzichtbar. Verbleibende mögliche Funktionen der Methylierungen des EWS-Proteins sind daher zum Beispiel die Modulation von Interaktionen zwischen RNA und dem EWS-Protein, die Steuerung der Aktivierungs-/Repressions-Aktivitäten des EWS-Proteins, die Stabilisierung des EWS-Proteins gegen proteolytische Degradierung oder aber als Quelle für asymmetrisch dimethyliertes Arginin, welches in der Zelle das Enzym Nitroxid-Synthase (NOS) reguliert. Da bis jetzt noch kein Enzym bekannt ist, welches eine Demethylierung katalysiert, ist eher eine statische Rolle der Methylierungen als eine regulatorische Funktion der EWS-Aktivität denkbar.

1. General Introduction

Post-Translational Modifications

In 1990, the human genome project was started with the goal to sequence all human chromosomes. This goal was achieved and it turned out, that the human organism has only 20000 to 25000 protein-coding genes (2). This rather small number of genes was unexpected, as the insect Drosophila melanogaster has already ~13500 genes and even the bacterium Escherichia coli with ~5000 genes has only a 4 fold smaller genome than human. An increase in complexity in higher eukaryotes is achieved at several levels of the protein biosynthesis. Various genes undergo alternative splicing after transcription, or different promoters for the same gene are used resulting in an estimated transcriptome of ~100000 human transcripts which is already a four- to five-fold increase in complexity. Another ten-fold increase in complexity is the result of post-translational modifications (PTMs). Once the protein is synthesized, manifold modifications like the cleavage of pro-peptides, the attachment of chemical groups (phospho-, acetyl-, or methyl groups, sugars, lipids, etc.) or internal and external cross-linking of polypeptide chains can direct the function of the protein. Taking such PTMs into account, the human proteome is by now estimated to contain more than 1000000 different proteins (3).

Therefore, the characterization of post-translational modifications and their responsible enzymes is necessary to understand the diversity of protein functions in the cell.

In this thesis, protein arginine methylations and its catalyzing enzymes were investigated in detail. Although this eukaryotic post-translational modification and its catalyzing enzymes are essential for the survival of the cell, it is hardly described in current biochemistry text books. The following chapter gives an overview about this in my opinion underestimated modification.

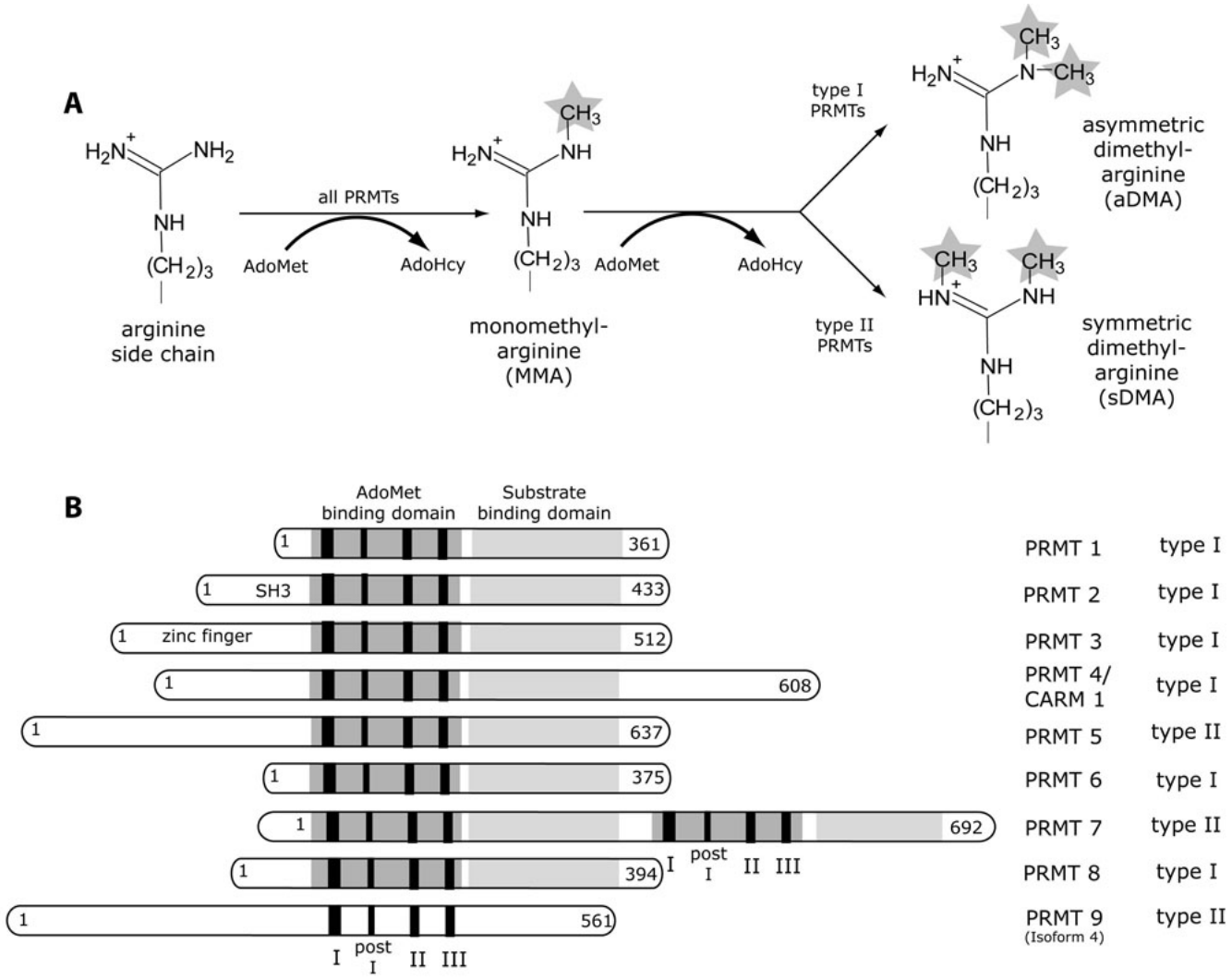


Figure 1.1: (**A)** Methylation of the arginine side chain by PRMTs. All PRMTs catalyze the formation of monomethyl-arginine (MMA), where the methyl donor S-adenosly-L-methionine (AdoMet) is converted to S-adenosyl-L-homocysteine (AdoHcy). In a second step, the type I PRMTs transfer a second methyl group to the same guanidino nitrogen resulting in an asymmetric dimethylarginine (ADMA) whereas the type II PRMTs catalyze the formation of symmetric dimethylarginines (SDMA). **(B)** Overview of the human PRMT family. The length of each protein is indicated by the numbers of amino acids. The PRMTs 1-8 all have a conserved AdoMet binding domain (shaded in grey) with the conserved motifs I, post I, II, and III, and a less conserved substrate binding domain (light grey). The conserved motifs I, post I, II, and III are present as well in PRMT 9, but the whole AdoMet binding domain shares only little homology to the other members of the PRMT family.

Protein Arginine Methylation

*(*summarized Review: "Steffen Pahlich, Rouzanna P. Zakaryan and Heinz Gehring. *Protein arginine methylation: Cellular functions and methods of analysis.* Biochim Biophys Acta. 2006 Dec;1764(12):1890-903.")

Arginine methylation is a common post-translational modification of mainly nuclear proteins in eukaryotic cells, and is catalyzed by a family of enzymes called protein arginine methyltransferases (PRMTs). S-Adenosyl-L-methionine (AdoMet) is used as the methyl donor in this reaction where the methyl group is transferred to one of the guanidinium nitrogens of arginine residues. PRMTs are classified into two groups. While type I PRMTs catalyze the formation of N^G-monomethylarginine (MMA) and asymmetric ω-N^G, N^G -dimethylarginine (ADMA), type II enzymes form MMA and symmetric ω- N^G, N'^G -dimethylarginine (SDMA,Fig 1.1 A).

Although arginine methylation was discovered more than 30 years ago (4,5), the responsible enzymes and the wide variety of substrates were largely unknown. In 1996, the first PRMT genes were cloned (6,7) and from then on, large quantities of pure and active recombinant enzyme was available for functional studies. Genetic screening yielded the identification of further PRMT genes and the knowledge and understanding of this modification increased steadily as indicated by the increase of scientific publications about arginine methylations from 1996 on (Fig. 1.2).

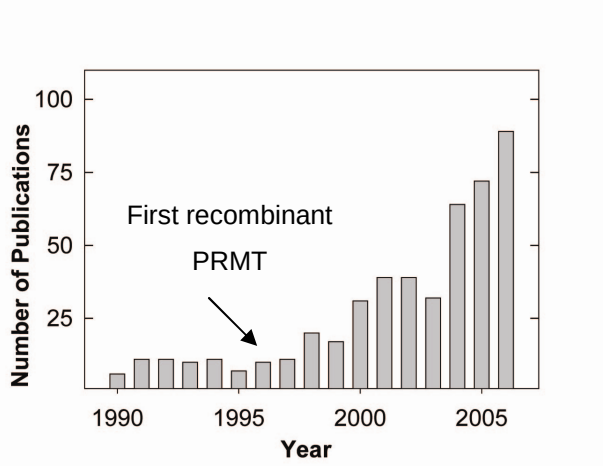


Figure 1.2: Scientific publications about arginine methylations from 1990 until 2006. The numbers were achieved using the PubMed database (http://www.ncbi.nlm.nih.gov/) and the search terms "arginine methylation" and the corresponding year.

Members of the PRMT family were found in many different eukaryotes from protozoa and fungi to higher plants and animals as highly conserved enzymes (8,9). Currently, nine human protein arginine methyltransferases are known: the type I methyltransferases PRMT1, PRMT3, PRMT4 (CARM1), PRMT6, and PRMT8, and the type II enzymes PRMT5, PRMT7, and

Table I: known substrates and intracellular localization of the PRMTs

PRMT, intranuclear localzation	Substrate
PRMT1 (type I), nuclear & cytoplasmic	53BP1
	CIRP
	EWS
	FGF-2
	Fibrillarin
	FKBP 12
	Fmrp
	GRP33
	GST-GAR
	GRY-RBP
	Hepatitis C virus NS3 Helicase
	Hepatitis delta antigen S-HDAg
	Histone H4
	hnRNP A1
	hnRNP A2
	hnRNP K
	hnRNP R
	ILF3
	Mre 11
	NF 90
	NIP45
	Nucleolin
	p137GP1
	PABP II
	PGC-1α
	QKI-5
	RBP58
	RNA helicase A
	SAF-A (hnRNP U)
	Sam68
	SAMT1
	SLM-1
	SLM-2
	SPT5
	TAFII68
	TIS21/BTG2
	TLS/FUS
	ZF5
HMT1 (yeast homolog), nuclear & cytoplasmic	Gar1p
	Hrp1
	Nab2p
	Nor1p
	Npl3
	Nsr1p
	Yra1
PRMT2 (type I), nuclear & cytoplasmic	not determined
HRMT1L1 (yeast homolog), nuclear	Adenovirus hnRNP E1B-AP5

PRMT, intranuclear localzation	Substrate
PRMT3 (type I), cytoplasmic	GST-GAR
	EWS
	Sam68
	PABP II
	hnRNP A1
	rpS2
PRMT4/ CARM1 (type I), nuclear	CARM1
	Histone H3
	PABP1
	p300/CBP
	TARPP
	ILF3
	HuR
	HuD
DART4 (Drosophila homolog), nuclear	Squid
	Vasa
PRMT5 (type II), nuclear & cytoplasmic	Histones H4, H2A, H3
	MBP
	Sm D1
	Sm D2
	SM B/B´
	coilin
	LSm4
	EBNA-2
	SPT5
PRMT6 (type I), nuclear	PRMT6
	GST-GAR
	Np13
	HIV Tat
	HMGA1a
	DNA polymerase beta
PRMT7 (type I), nuclear & cytoplasmic	GST-GAR
PRMT8 (type I), membrane-associated	GST-GAR
	GST-Npl3
	Histone H4
PRMT9 (type II), nuclear & cytoplasmic	not determined
not determined	Vav-1
not determined	Myosin
not determined	Herpes simplex virus ICP27
not determined	RBP16
not determined	Golgi proteins

PRMT9. PRMT2 was identified by sequence homology, however, as yet no methyltransferase activity could be determined (10) (Table I). Some homologues from other species are also listed in the table. Crystal structure analysis revealed a strong structural conservation of the AdoMet-binding site of different PRMTs (for reviews, see (11,12)) which however differ in N- and C-terminal extensions (Fig. 1.1 B). PRMT9 structurally differs from all other known PRMTs, implying convergent evolution of enzymic mechanism of arginine methylation (13).

PRMT1, the main methyltransferase in human cells with a wide substrate spectrum, is responsible for more than 80% of cellular PRMT activity (14). It generally recognizes arginines within a glycine-arginine-rich (GAR) region, a motif which is present in many RNA- or DNA-binding proteins. The other type I enzymes PRMT3, 6, and 8 also recognize GAR motifs whereas for PRMT4/CARM1 no preferred methylation motif is known. The type II enzymes PRMT5 and 7 methylate GAR motifs as well as arginines situated not in a known consensus region. Many methylated proteins have been identified to date, but the responsible methyltransferases are still largely unknown. The known substrates as well as the different subcellular localizations of the PRMTs are summarized in Table I.

Arginine methylation and protein function

Arginine methylation is implicated in many cellular processes including transcription, RNA processing and transport, translation, signal transduction, DNA repair, apoptosis etc. In the following, the main functions are summarized.

Transcriptional regulation

Methylation of nucleosomal histones by PRMTs plays an important and dynamic role in nuclear receptor-mediated transcriptional regulation, chromatin remodelling and other aspects of gene expression (reviewed in (11,15)). Linkage of histone arginine methylation, lysine methylation, acetylation, phosphorylation and ubiquitination, regulates binding of specific interacting proteins, altering chromatin structure and transcription process (16).

Arginine methylation participates also in the regulation of the initiation and elongation steps of transcription. The methylation of the histone acetyltransferase CREB binding protein (CBP), which functions with the related p300 protein as a key transcriptional coactivator, by CARM1 blocks its CREB activation (17,18). PRMT1 and PRMT5 have been shown to methylate the transcriptional elongation factor SPT5 and thus to regulate its interaction with

RNA polymerase II (19), and arginine methylation by PRMT6 inhibits the transactivation activity of HIV Tat protein, therefore PRMT6 acts as a restriction factor for HIV replication (20).

RNA processing and export; protein translation

RNA-binding proteins (RBPs) participate in processing, folding, stabilization and localization of RNAs and mRNA translation. Most of RBPs possess GAR motifs, and are well-known or putative substrates for PRMTs. Arginine residues in the binding sites of RBPs have been recognized as key amino acids in RNA-protein interactions and its methylation can facilitate as well as repress the interaction with RNA.

It was demonstrated in yeast, that methylation activity is needed for the recruitment of mRNA-export factors during transcriptional elongation (21) and ribosome biosynthesis (22). Methylation of the RNA-binding proteins HuD (23) and HuR (24) by CARM1 has been shown to regulate their function in stabilization of specific mRNAs which is important in cell proliferation and differentiation. Arginine methylation of several RNA-binding proteins is needed for their proper subcellular (nuclear) localization (25,26) as demonstrated for SAM68 (27), hnRNP A2 (28), or hnRNP Q (29).

Signal transduction

Cell signaling is controlled by posttranslational modifications that alter protein-protein interactions. More and more data show that also arginine methylation is used to mediate signal transduction downstream of the TCR (T-cell antigen receptor), cytokine (including interferon), and NGF receptors (for reviews, see (8,30)).

DNA repair

53BP1, a central mediator of the DNA damage checkpoint, is rapidly recruited to sites of DNA double-strand breaks. Whereas oligomerization of 53BP1 was shown to be independent of methylation (31), asymmetric dimethylation of the GAR motif of 53BP1 by PRMT1 was demonstrated to be required for DNA binding activity (32). PRMT6 was found to play a direct role in the regulation of DNA repair processes by methylation of DNA polymerase β, which participates in base excision repair (33).

Protein-protein interactions

Arginine methylation can increase or decrease protein-protein interactions. The methylation of an arginine residue of the HIV-1 Nef protein or hnRNP K inhibits their interaction with tyrosine kinases Fyn and c-Src, respectively (34), (35), and methylation of Sam68 prevents its association with SH3 domains. On the other hand, the interaction between the nuclear factor of activated T cells (NF-AT) and the NF-AT-interacting protein of 45 kD, NIP45, decreased in the presence of methylase inhibitors ((8) and references therein).

Arginine protection

In addition to the above mentioned functions, it was proposed that arginine methylation might protect crucial arginine residues against attack by endogenous reactive dicarbonyl agents, such as methylglyoxal and other natural by-products of normal metabolic pathways (36). An experimental confirmation, however, is still missing.

Arginine methylation as a source of free MMA, ADMA and SDMA

Proteolysis of methylated proteins is the cellular source of free mono- and dimethylarginine residues. MMA and ADMA are known to be endogenous competitive inhibitors of nitric oxide synthase (NOS), which synthesizes NO from L-arginine, and plays an important role in cardiovascular system. Dimethylarginine dimethylaminohydrolases (DDAHs) specifically hydrolyze MMA and ADMA, but not SDMA. Deregulated levels of these free methylarginines affects NOS activity and, consequently, may result in cardiovascular disease ((37) and references therein).

Demodification of methylated arginines

Over the years, arginine methylation has been considered to be an irreversible posttranslational modification. The search for arginine demethylases did not yield any results. Nevertheless, the fact that lysine demethylases such as LSD1 have been identified makes it possible that arginine demethylases may also exist. Two groups showed that arginine methylation of histones can be reduced in vitro by deimination with PAD4 (38,39). This reaction converts both unmodified arginine and monomethylarginine, but not dimethylarginine, to citrulline, an amino acid that is not transcriptionally incorporated into proteins. However, dimethylated arginines, which is the dominant methylated amino acid in a cell extract (~70%), can not be converted to citrulline, and enzymes, catalyzing the conversion

from citrulline back to arginine are also unknown. Therefore, deimination is rather the exception than the rule in arginine demethylation.

Methods to detect and identify arginine methylated proteins and their methylation sites

In vitro methylation with [^{3}H]-AdoMet

A common way to analyze arginine methylation is the in vitro methylation by using recombinant methyltransferases, radioactively labeled S-adenosyl-L-[methyl-^{3}H]-methionine ([^{3}H]-AdoMet), and purified proteins or cell extracts as substrate. The labeled methyl groups are transferred to the substrates and can be detected by fluorography (40). However, in eukaryotic cells most methyl-accepting sites are occupied indicating that methylation occurs soon after translation (41,42) and untreated cell extracts are not suitable as substrate. To obtain free methylation sites of methyl accepting proteins, cellular protein methyltransferases are usually inhibited by adenosine dialdehyde (AdOx). AdOx inhibits S-adenosyl-L-homocysteine hydrolase, leading to an accumulation of intracellular S-adenosyl-L-homocysteine, which in turn is a product inhibitor of the methyltransferases (43). Cells treated with AdOx accumulate newly synthesized unmethylated methyl-accepting proteins during the incubation period (42,44) and can be used as substrate for in vitro methylations. However, one has to consider that the expression level of originally methylated proteins might be affected by the inhibitor treatment (45). Due to the incorporation of the labeled methyl group, methylation activity of the used enzyme can be detected by fluorography after separation of the proteins by SDS-PAGE. Using hypomethylated cell extracts, different substrates as well as substrate specificities of the PRMTs were observed with this method (46-50). It is easy and rapid to detect or proof methyltransferase activity of PRMTs against single substrate proteins or hypomethylated cell extracts. However, when hypomethylated cell extract is used as substrate, an identification of unknown methylated proteins is difficult and no information about the position of the methylated arginines within the proteins can be gained.

Identification of the type of methylation by amino acid analysis

The type of methylation, namely MMA, ADMA, or SDMA, can be distinguished by amino acid analysis of the methylated sample. The methylated proteins or peptides have to be hydrolyzed completely and the resulting amino acids have to be separated

chromatographically (51,52). MMA, ADMA, and SDMA differ in their retention times or Rf-values and can be assigned with the corresponding standards.

Detection of methylated arginines by antibodies

The different types of arginine methylation can also be analyzed by Western blotting using methyl-arginine-specific antibodies. This technique was used to identify the asymmetric dimethylation of the histones H3 and H4 (53,54), or SAM68 (27), the symmetric dimethylation of coilin (55), or the conversion of unmodified or monomethylated arginines in H3 and H4 to citrulline by PAD4 (38,39).

Commercially available antibodies (SYM10, SYM11, ASYM24 from Upstate or 7E6 from Abcam) are derived from synthetic peptides of several RG-repeats including either SDMA or ADMA (55) or as in the case of histones from a peptide covering the exact sequence including the modified arginine (38). These antibodies do not necessarily recognize all and only methylated proteins. With the 7E6-antibody, cross-reactivities could not be excluded (42), and also the specific recognition of methylated STAT1 and STAT3 by antibodies is questionable (56). So for each analysis, the specificity of the used antibodies has to be tested.

Detection of methylated arginines by mass spectrometry (MS)

MS has become a favorable method not only to identify proteins by peptide mass fingerprint, but also to characterize post-translational modifications of proteins. Methylation leads to a mass increase of 14 Da per incorporated methyl group at the arginine residue. After enzymic digestion of a methylated protein, the mass shift can be used to find peptides that potentially contain this modification (Fig. 1.3) either in endogenously methylated proteins purified from a cell extract (57,58) or after in vitro methylation of recombinant unmodified proteins with different PRMTs (59,60). Arginine methylation does not change the ionization properties of the modified peptide and therefore, the detection of unmodified and modified peptides by MADLI-TOF-MS is not influenced by this modification. To confirm and detect the methylated amino acid, the corresponding peptide can be sequenced by MSMS. A specific fragment ion at m/z 46.06 corresponding to dimethylammonium is exclusively generated from the fragmentation of ADMA-containing peptides and can be thus used as a specific reporter ion for the detection of ADMA (61,62).

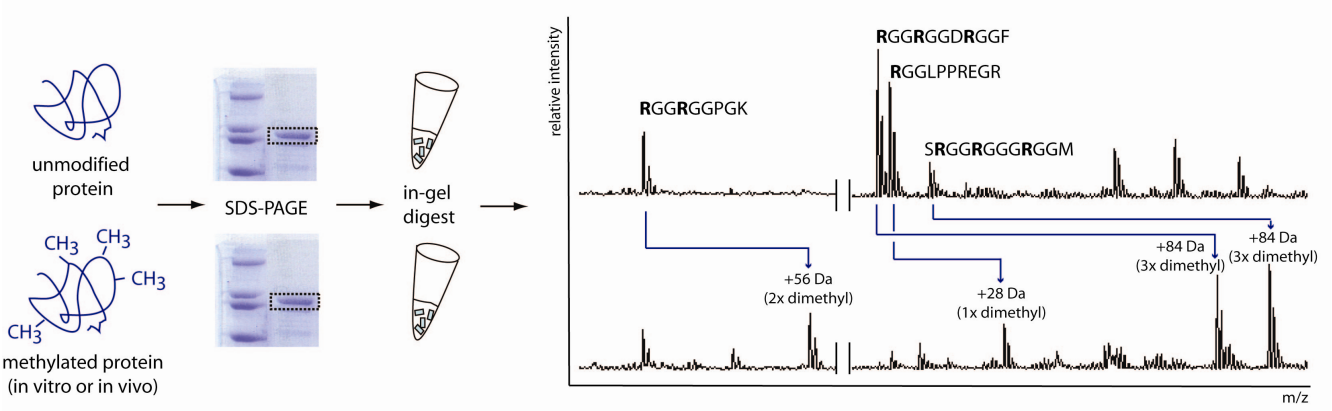


Figure 1.3: Detection of protein arginine methylation by peptide mass fingerprints. Potential methylation sites of a protein can be identified by peptide mass fingerprint. Here, recombinant (unmethylated) human EWS protein was in-gel digested and analyzed by MALDI-TOF MS. The same was done with recombinant EWS protein, prior methylated with PRMT1. Methylation leads to a mass shift of multiples of 14 Da of peptides containing potential methylation sites like RGG motifs.

Proteome analysis of arginine methylated proteins

The methods described above are mainly suitable for the characterization or identification of single proteins and their modifications. For a proteomic approach, however, the whole proteome or subproteomes like the phosphoproteome (selective enrichment of phosphorylated proteins) of an organism or cell type needs to be analyzed. Enrichment and analysis of a "methyl-proteome", however, is difficult because the incorporation of methyl groups to the arginine side chain changes the physicochemical properties of this residue only slightly. Although the hydrophobicity is enhanced, the charge remains unchanged and methylated proteins or peptides can not be separated, so far, from unmodified ones by any chromatographic method as e.g. phosphoprotein or -peptide enrichment by immobilized metal affinity chromatography (IMAC).

In the following, methods are described to enrich or visualize methylated proteins for a "methyl-proteome" analysis.

In vivo methylation with [^{3}H]-AdoMet

Yeast cells are capable to take up the methyl donor AdoMet from the medium (63). Proteins from yeast cells, grown in [^{3}H]-AdoMet-containing medium were separated by 1D or 2D gel electrophoresis and the methylated proteins were detected by fluorography (64). AdoMet-based methylation of lysines, histidines, or nucleic acids leads also to radioactive labeling. In rat cells, however, it has been shown that almost 90% of methylated amino acids of the cell

extract are mono- and asymmetrically dimethylated arginines (65). If one considers a similar proportion of methylated amino acids in yeast, methylation of other residues than arginines is expected to be a minor problem. The in vivo labeling, however, is limited to the analysis of protein methylation in yeast, because most other eukaryotic cells have only a poor uptake efficiency of AdoMet (66).

Instead, Najbauer et al. (65) used a hypomethylated cell extract of PC12 rat cells to visualize methyl acceptor proteins. PC12 cells were incubated in culture prior to lysis with AdOx, and [^{3}H]-AdoMet was then added to the cell extract to methylate the proteins by endogenous methyltransferases present in the extract. The proteins were separated by 2D gel electrophoresis and more than 50 different methyl acceptor proteins, mainly basic proteins, could be detected by fluorography. However, an identification of these proteins was not possible and methylation of hypomethylated cell extracts might not reflect the "methyl-proteome" of normally grown cells.

To avoid hypomethylation, cells can be incubated in methionine-free medium containing labeled methionine in the presence of protein synthesis inhibitors like chloramphenicol and cycloheximide. The labeled methionine is taken up by the cells and converted intracellularly into the labeled methyl-donor AdoMet. The incorporation of the labeled methyl group to a protein is due to the AdoMet-based methylation process and not due to incorporation of labeled methionine into newly synthesized proteins. With this method, Liu et al. (67) discovered the methylation of numerous heterogeneous nuclear ribonucleoproteins (hnRNPs) by immunopurification and 2D gel electrophoresis of hnRNP complexes and detection of the labeled methylated hnRNPs by fluorography.

Immunopurification of methylated proteins

Boisvert et al. (68) used antibodies derived against synthetic peptides containing either SDMA or ADMA in different RG-rich motifs (see above) to immunoprecipitate methylated proteins out of a HeLa cell extract. The precipitated protein complexes were digested and more than 200 proteins were identified by LC/MSMS sequencing of the peptides. Proteins were assumed to have methylated arginines if they contain GAR-motifs within their amino acid sequence but were absent in a control experiment. Peptides containing methylated arginine residues, however, were not directly identified. Discrimination between potential asymmetric or symmetric dimethylated proteins was based on the used antibodies and discrimination between methylated and co-purified unmethylated proteins was not possible.

Ong et al. (69) introduced heavy methyl SILAC (stable isotope labeling by amino acids in cell culture) to identify methylated proteins from cell extracts. Cells were grown in medium containing either "light" (CH_3)- or "heavy" ($^{13}CD_3$)-methionine, which is taken up by the cells and metabolically converted to labeled AdoMet, used for intracellular methylation events. Both samples were combined in a 1:1 ratio, subjected to a single immunopurification step with specific antibodies (see above), digested, and analyzed by high mass accuracy LC/MSMS. The incorporation of one "heavy" methyl group leads to a mass increase of 4 Da compared to the "light" one. Thus, ion pairs with equal intensity ratios and distances corresponding to the incorporation of one or multiple methyl groups (e.g. 4 Da for MMA, 8 Da for DMA) are hints for methylation. Peptides with such ion pairs were sequenced by MSMS to identify the position of the methylated arginine residues. With this promising method, both, identification and relative quantification of the methylation site are feasible. However, to reduce sample complexity, an immunopurification step with antibodies against methylated arginines was performed, and only 33 methylated proteins out of a whole HeLa cell extract were identified in this study. This modest enrichment of methylated proteins again reflects the moderate specificity of currently available antibodies against methylarginines for immunoprecipitation of the "methyl-proteome".

In summary, a variety of methods for the analysis of protein arginine methylation are currently available, especially for in vitro studies. A complete "methyl-proteome" analysis can reveal the number of methylated proteins in the cell and give an idea of the variety of cellular processes this modification is involved in, but this approach remains a challenge. High throughput mass spectrometry makes it possible to identify huge amounts of different proteins and peptides in a short time, but there is still a lack of methods in efficiently enriching methylated proteins out of a whole proteome.

The Ewing Sarcoma (EWS) Protein

The human Ewing sarcoma (EWS) protein, a multifunctional RNA-binding protein, was found in previous studies to contain multiple asymmetrically dimethylated arginine residues within its RNA-binding domain (RBD) (57). The functional role of this modification as well as the responsible methyltransferases, not known so far, was investigated in this study. In the following, the current knowledge about the structure and function of the EWS protein is summarized.

EWS protein and cancer

The human *EWS* gene is located at the chromosome 22 (22q12) and codes for the 656 amino acid long Ewings sarcoma (EWS) protein. In Ewing sarcoma family of tumors (ESFT) a part of the *EWS* gene is translocated resulting in the expression of fusion proteins between the N-terminal transcriptional activation domain (EAD) of the EWS protein and the DNA-binding domain of one of several members of the ETS (erythroblastostosis virus-transformed sequence) transcription factors (70). ESFT are the second most malignant neoplasm of bone among children and adolescents. The most frequent chromosomal translocation t(11;22)(q24;q12) results in the expression of the oncoprotein EWS/Fli1 (90-95% within ESFT) (71), but also other cancer-associated gene fusions involving the EWS gene, like EWS/ATF1 (malignant melanoma of soft parts) (72), EWS/WT1 (desmoplasmic small round-cell tumors) (73,74), or EWS/CHOP (myxoid liposarcoma) (75) are observed. In all cases, the fusion proteins act as transcriptional activator, modulating the expression of target genes different from those recognized by the proto-oncoproteins (70), but the detailed mechanism of transformation by EWS fusion proteins is still open.

In the wild-type EWS protein, the transforming potential of the EAD is repressed by the C-terminal RNA-binding domain (RBD) (76), but the mechanism and the cellular function of the proto-oncoprotein EWS in general is largely unknown. The determination of the cellular role of the wild type EWS protein will therefore contribute to the understanding of the tumor-associated transformation by EWS fusion proteins.

Structure of the EWS protein

The EWS protein consists of two functional domains, an N-terminal transcriptional activation domain (EAD) and a C-terminal RNA-binding domain (RBD,Fig. 1.4).

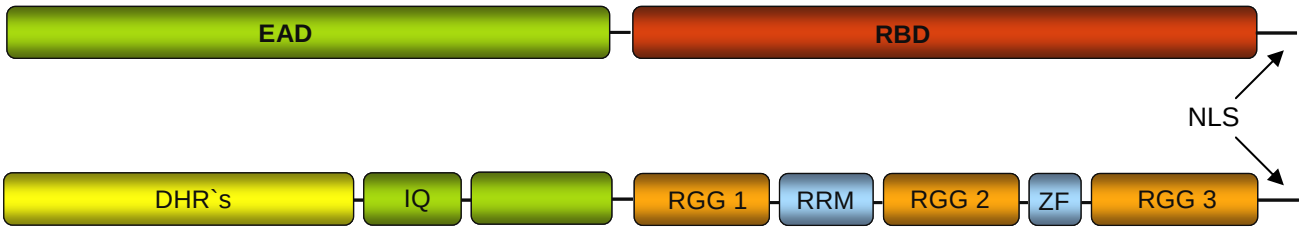


Figure 1.4: Structural and functional motifs of the EWS protein. The abbreviations used are: EAD, N-terminal transcriptional activation domain; RBD, RNA binding domain, DHR`s, degenerated hexarepreats with the consensus sequence SYGQQS; IQ, IQ domain which binds calmodulin; RGG, arginine-glycine rich regions; RRM, RNA-recognition motif; ZF, putative C_2C_2 zinc finger motif; NLS, nuclear localization sequence.

N-terminal transcriptional activation domain (EAD)

The N-terminal transcriptional activation domain (EAD, AA 1-299) of the EWS protein is composed of multiple degenerated hexapeptide repeats (DHR) with the consensus sequence SYGQQS including a highly conserved Tyr residue. The EAD of the EWS protein does not possess any secondary or ternary structure and due to its amino acid composition, it is most likely an intrinsic disordered domain. A prediction of intrinsically disordered regions using PONDR (Predictor of Naturally Disordered Regions, www.pondr.com) revealed ~80 % of the EWS protein to be disordered, including the whole EAD (77). The amino acids 258-280 build an IQ domain, which binds calmodulin and can be phosphorylated by protein kinase C (78).

C-terminal RNA-binding domain (RBD)

The C-terminal RNA-binding domain (RBD, AA 300-656) of the EWS protein consists of an RNA-recognition motif (RRM, AA 361-447) and a putative C_2C_2 zinc finger (AA524-543), surrounded by arginine-glycine rich regions (RGG-box 1-3). Finally, the last 17 amino acids define a new nuclear localization signal (C-NLS), as recently discovered in our lab (79) and elsewhere (80).

RNA-recognition motif (RRM)

RNA-recognition motifs (RRM) are well known domains of RNA- or DNA-binding proteins. This motif is one of the most abundant protein domains in eukaryotic proteins. By now, approximately 6000 RRMs were found in ~ 3500 different proteins, and 2% of the

human gene products contain one or several RRMs. Recently, RRMs were also identified in prokaryotic proteins (85 proteins in bacteria, mainly cyanobacteria, and six proteins in viruses), but their roles in bacteria and viruses is still unknown (81).

RRMs consist of an αβ-sandwich structure with a $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ topology. The antiparallel β-sheets form a flat surface which serves as docking platform for RNA or DNA. The two α-helices are packed against the β-sheets whereas the N- and C-terminal regions of the RRM are unstructured linker. The interaction between the RRM and nucleic acids is maintained by three highly conserved aromatic amino acids (Phe or Tyr) within the β-sheets 1 and 3, which interact hydrophobically with the bases of the nucleic acids. The high affinity and sequence specificity of different RRMs towards their RNA- or DNA targets is achieved by the assistance of the external β-sheets 2 and 4, the loops between the β-sheets and the α-helices, or by the association of multiple RRMs (81).

The structure of the EWS-RRM was solved by NMR (Nagata et al. 2005, Research Collaboratory for Structural Bioinformatics (RCSB) database (www.rcsb.org), unpublished data) and shows the typical β-sheet platform and the two α-helices (Fig. 1.5 A). Interestingly, a sequence alignment with the RNA-binding protein hnRNP A1 shows, that the two aromatic amino acids in β-sheet 3, responsible for RNA-binding via hydrophobic interaction, are replaced in the EWS RRM (and its homologues TLS/FUS and $TAF_{II}68$) by a threonine and an aspartic acid (Fig. 1.5 A). In vitro RNA-binding activity of the EWS protein was demonstrated towards poly G- and poly U-RNA via its C-terminal RGG box (82), however, no direct binding of RNA or DNA to the EWS RRM was demonstrated yet. RRMs are also known to mediate protein-protein interaction, as it was demonstrated for the exon-junction complex protein Y14, which interacts via its β-sheet surface of the RRM with magoh (83). The RRM of the EWS homolog TLS was found to be dispensable for RNA binding in vivo (84), and in vitro (85). Thus, it is reasonable, that the RRM of the EWS protein (and its homologues) might serve to maintain rather protein-protein than protein-RNA interactions.

Zinc finger motif (ZF)

The RBD of the EWS protein contains a putative C_2 C_2 zinc finger motif, also called RanBP zinc finger due to its homology to the nuclear export protein RanBP2 zinc finger. Such zinc finger motifs sandwich a single zinc atom by two short β-hairpins. The RanBP zinc finger of ZNF265 was found to bind mRNA (1), the one of Npl4 interacts with ubiquitin (86), and the TLS/FUS zinc finger is able to bind a GU-rich RNA oligomer (85). The functional role as

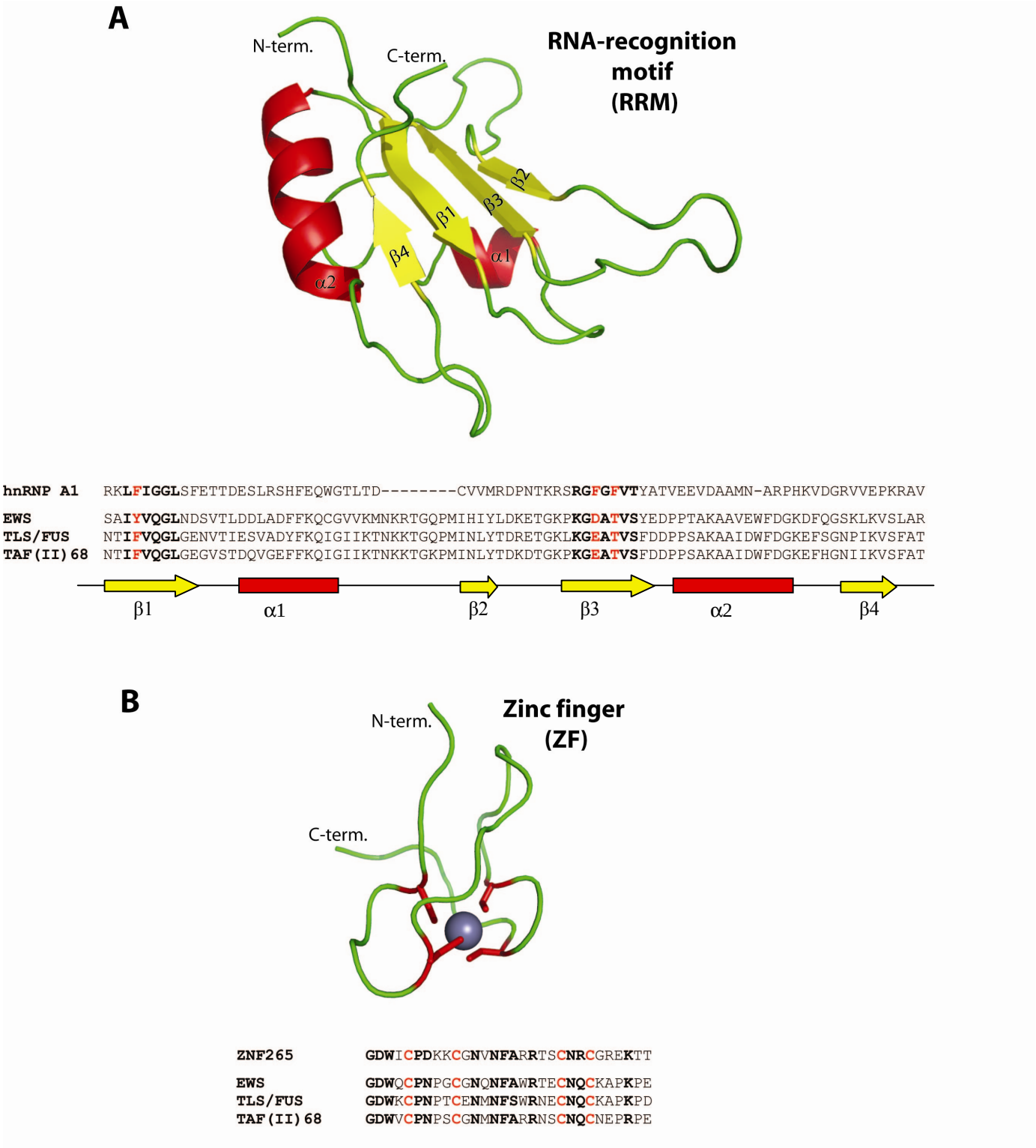


Figure 1.5: (A) NMR structure of the RNA-recognition motif of the EWS protein (Nagata et al., 2005, unpublished data). The sequence alignment of the hnRNP A1 RRM and the RRMs of the TET family proteins was done with ClustalW. The three aromatic amino acids of hnRNP A1 within β1 and β3 labeled in red are necessary for RNA interaction. In the β3 of the TET-family proteins EWS, TLS, and $TAF_{II}68$, the aromatic amino acids are replaced by aspartic (or glutamic) acid and threonine. **(B)** Homology model of the EWS zinc finger structure based on the known NMR structure of ZNF265 (1). The four cystein side chains labeled in red are chelating the zinc atom (grey). Sequence alignment of ZNF265 and the TET family proteins was done with ClustalW. Both structures were visualized with the program PYMOL.

well as the structure of the EWS zinc finger is not solved so far, but the high sequence similarity to the ZNF256 zinc finger, whose structure was solved by NMR (1), allows homology modeling to get the structure of the EWS zinc finger (Fig 1.5 B), as it was done for TLS/FUS (85). For ubiquitin-binding of the Npl4 zinc finger, a Thr-Phe dipeptide between the two cysteins is needed. In case of EWS and TLS/FUS, this dipeptide is replaced by Glu-Asn, so the EWS zinc finger might rather functions in RNA-interaction than contributing to ubiquitin-binding

Arginine-glycine rich regions (RGG boxes)

The two structured domains of the EWS RBD (RRM and ZF) are surrounded by arginine-glycine rich domains (RGG boxes 1-3). Such RGG boxes are found in many RNA-binding proteins like heterogeneous nuclear ribonucleoproteins, nucleolin, fibrillarin, or the EWS homologues TLS/FUS and $TAF_{II}68$. These RGG- or RGR-repeats are the preferred substrates for protein arginine methylation, and in the EWS protein all 30 arginines within the RGG boxes were found to be dimethylated in vivo (57). The functional role of the EWS protein RGG boxes and their methylation is still unknown. RGG- boxes, unmethylated or methylated, do not possess any stable secondary or ternary structure and seem to be flexible linker between structured domains (85). The RGG rich region of the fragile X mental retardation protein (FMRP) specifically binds mRNA (87). Because recombinant produced peptides, which can not be methylated in bacteria, were used in this study, the role of FMRP-methylation in mRNA-binding needs to be investigated.

Nuclear localization sequence (NLS)

The last 17 C-terminal amino acids (639-656) of the EWS protein form a new nuclear localization sequence. The functional key amino acids within the NLS are Arg 648, Arg 652, Pro 655, and Tyr 656, as the mutation of each of these residues to Ala prevents the nuclear import of the EWS protein (79). The so called PY-NLS, also present in hnRNP A1, was found to bind to Karyopherinβ2 (Transportin) and currently, further 81 proteins were predicted to contain such a PY-NLS (80).

Subcellular localization of the EWS protein

The EWS protein was identified by immunostaining and overexpression as a GFP-fusion protein to localize mainly in the nucleus of several cell lines (79). Differential centrifugation and surface labeling of cells revealed that subpopulations of the EWS protein are also located

in the cytoplasm and attached to the cell membrane (57,45). Other nuclear proteins like nucleolin and hnRNP K are known to locate additionally at the cell surface and a receptor function of these proteins was proposed (88,89). However, receptor activity or specific ligands of the surface-exposed EWS protein are not known so far.

Post-translational modifications of the EWS protein

The EWS protein is heavily post-translationally modified. All 30 arginines within the RGG boxes were found to be asymmetrically dimethylated, independent on the isolation of the EWS protein from the nuclear or cell-membrane fraction (57). By now, no unmethylated EWS protein was isolated from cells. Arginine methylation can affect subcellular localization as demonstrated for hnRNP A2 (28) and hnRNP Q (29) and a similar function was proposed for the EWS protein. Recently, we could show in our lab, that the subcellular localization of the EWS protein is not affected by this modification (79). Inhibition of the methyltransferases in the cell results in a decreased expression level of the EWS protein (45).

Additionally to the arginine methylation, the EWS protein is supposed to be phosphorylated. The Ser266 was shown to be phosphorylated by protein kinase C, which prevents calmodulin binding to the IQ domain (78) and regulates the transcriptional activity of the fusion proteins EWS/Fli1 and EWS/Atf1 (90). Several Ser and Tyr residues of the EAD are predicted to be phosphorylated, however, an experimental confirmation is still missing. It is known, that especially disordered regions are the target for numerous post-translational modifications as it was shown e.g. for p53 (91,92), c-Src kinase (93), or the unstructured N-terminal tail of histones, whose phosphorylation, acetylation, and methylation is crucial for the histone functions (94). Consistent with that, only the disordered parts of the EWS protein are modified, whereas the structured RRM and ZF domains harbor no PTMs.

Cellular functions of the EWS protein

Transcriptional co-activation

The N-terminal EAD in oncogenic EWS fusion proteins function as potent transcriptional activator. Although the transforming potential is repressed in the wild-type EWS, it still acts as a transcriptional (co)-activator. The EWS protein associates with integral components of the transcriptional pre-initiation complex like the C-terminal domain of the RNA polymerase II, and the TFIID complex (95). A direct function as co-transcriptional activator was shown

by the interaction of the EWS protein with the transcriptional activators CBP/p300 and their co-activation in a cell line- and promoter-specific manner (96). The EWS protein also interacts with Brn-3a, a POU-family transcription factor involved in the regulation of genes that control neural development. Additionally, the transcription of a Brn-3a isoform is activated by the EWS protein (97). Oct-4, a transcription factor which also belongs to the POU family, is expressed in embryonic stem cells and activates genes involved in maintaining an undifferentiated totipotent state by preventing differentiation. The EWS protein was found to interact with the POU domain of Oct-4 and to increase its transcriptional activity (98).

As already mentioned the wild type EWS protein (co)-activates different genes and transcription factors than the chimeric EWS fusion proteins. Alex et al (2005) could show, that multiple RGG boxes of the C-terminal RBD of the EWS protein, which are absent in the fusion proteins, can repress the transcriptional activity of EWS fusion proteins. Also other well known activation domains like CREB or VP16 were repressed when fused to the RGG boxes, but the presence of more than one activation domains diminished the repression (76). The C-terminal RBD of the EWS protein including the RGG boxes therefore plays an important role in the regulation of its N-terminal activation domain. Interestingly, the replacement of the arginine residues by lysines did not alter the EAD repression (76), so the basic charge and not the methylation of the arginines seem to be important for repression.

Splicing

A role of the EWS protein in splicing was proposed due to its interaction with several splicing factors. By yeast-two hybrid screens it was found, that the EAD of the EWS protein interact with the splicing factors U1C (99) and ZFM1 (100). The RBD can also interact with Ser/Arg-rich splicing factors TASR-1 and TASR-2 (101,102) and with the splicing factor YB-1 (103). The EWS protein co-purifies with the heterogeneous nuclear ribonucleoproteins (hnRNP) A1 and C1/C2 (104) and in a large scale proteomics chacacterization of the human spliceosome, more than 300 different components of this huge and highly dynamic protein complex were identified, among them the EWS protein (105). The contribution of the EWS protein in pre-mRNA processing and splicing, however, is still unknown.

mRNA transport and translation

Splicing and m-RNA export out of the nucleus are highly related processes. Many RNA-binding proteins, like hnRNP A1 were found to bind already to the pre-mRNA and remain bound until the spliced mRNA product is exported to the cytoplasm and translation is initiated.

Other proteins like poly(A)-binding protein (PABP)-2 dissociate in the nucleus from the protein-RNA complex, whereas new binding partners like PABP-1 are recruited in the cytosol.

Usually, the EWS protein is located in the nucleus and is not supposed to frequently shuttle between nucleus and cytosol. Nevertheless, the EWS protein might contribute in the transport of spliced mRNA, as it was identified in RNA-transporting granules in neurons, which move along the microtubules mediated by kinesin (106). Angenstein et al. investigated the protein composition of messenger ribonucleoprotein (mRNP) complexes bound either to non-translated or to translated poly(A)-mRNA. The EWS protein was identified only in the non-translated fraction and not in the polysome-bound mRNPs indicating, that it does not play a direct role in translation (107).

DNA-recombination, -pairing, and -repair

The three TET-family proteins TLS/FUS, EWS, and TAFII68 have DNA-pairing activity as it was shown by Guipaud et al. (108). Consistent with these results, TLS-deficient mice fail to develop B-lymphocytes, a process which is characterized by a couple of gene rearrangements. Recently, Li et al. succeeded to produce also EWS null mutant mice, which show a very high percentage of mortality (90%). Similarly, the few surviving EWS -/- mice fail to develop intact B-lymphocytes. Whereas the lack of TLS results in B-cells with dysfunction in the primary antibody production (109), the EWS -/- mutants have a defect in pre-B-lymphocyte development during pro- to pre-B cell transition (110). The deletion of the EWS gene affects also meiosis, as the formation of bivalent XY chromosomes is drastically reduced (110). This defect in formation of bivalent chromosomes during meiosis was also observed in TLS -/- mutants, however, there the pairing of the autosomal chromosomes was affected (111). Finally, EWS-deficient cells are highly sensitive to radiation, implying also a function in DNA repair (110).

Signal transduction

A role in signal transduction was proposed, as the EWS protein interacts with calmodulin via its IQ domain. This interaction can be repressed by phosphorylation of Ser266 by protein kinase C (PKC) (78). Also the SH3 domain of Brutons kinase (112) and Tyrosine kinase Pyk2 were found to interact with the EWS protein, the latter one is supposed to function in G-coupled receptor signaling (113).

In summary, the EWS protein is an intrinsic disordered protein with only two structured domains (RRM and ZF). It is extensively methylated at arginine residues and supposed to be phosphorylated manifold. Beside the cancer-associated transactivation activity of the fusion protein, the wild-type EWS protein participates in several cellular processes like transcriptional control, splicing and mRNA processing, DNA-pairing and –repair, and signal transduction. Such multi-functionality is characteristic for intrinsic disordered proteins, as the disordered domains enable a high binding diversity by interaction with numerous binding partners with high specificity and low affinity. However, the role of methylation and phosphorylation in all these processes the EWS protein is involved in has not been addressed.

References

1. Plambeck, C. A., Kwan, A. H., Adams, D. J., Westman, B. J., van der Weyden, L., Medcalf, R. L., Morris, B. J., and Mackay, J. P. (2003) *J Biol Chem* 278(25), 22805-22811
2. Consortium, I. H. G. S. (2004) *Nature* 431(7011), 931-945
3. Jensen, O. N. (2004) *Current opinion in chemical biology* 8(1), 33-41
4. Paik, W. K., and Kim, S. (1967) *Biochem Biophys Res Commun* 29(1), 14-20
5. Paik, W. K., Paik, D. C., and Kim, S. (2007) *Trends in biochemical sciences* 32(3), 146-152
6. Gary, J. D., Lin, W. J., Yang, M. C., Herschman, H. R., and Clarke, S. (1996) *J Biol Chem* 271(21), 12585-12594
7. Henry, M. F., and Silver, P. A. (1996) *Molecular and cellular biology* 16(7), 3668-3678
8. Boisvert, F. M., Chenard, C. A., and Richard, S. (2005) *Sci STKE* 2005(271), re2
9. Miranda, T. B., Sayegh, J., Frankel, A., Katz, J. E., Miranda, M., and Clarke, S. (2006) *Biochem J* 395(3), 563-570
10. Scott, H. S., Antonarakis, S. E., Lalioti, M. D., Rossier, C., Silver, P. A., and Henry, M. F. (1998) *Genomics* 48(3), 330-340
11. Lee, D. Y., Teyssier, C., Strahl, B. D., and Stallcup, M. R. (2005) *Endocr Rev* 26(2), 147-170
12. Cheng, X., Collins, R. E., and Zhang, X. (2005) *Annu Rev Biophys Biomol Struct* 34, 267-294
13. Cook, J. R., Lee, J. H., Yang, Z. H., Krause, C. D., Herth, N., Hoffmann, R., and Pestka, S. (2006) *Biochem Biophys Res Commun* 342(2), 472-481
14. Tang, J., Kao, P. N., and Herschman, H. R. (2000) *J Biol Chem* 275(26), 19866-19876
15. Wysocka, J., Allis, C. D., and Coonrod, S. (2006) *Front Biosci* 11, 344-355
16. Zhang, Y., and Reinberg, D. (2001) *Genes Dev* 15(18), 2343-2360
17. Xu, W., Chen, H., Du, K., Asahara, H., Tini, M., Emerson, B. M., Montminy, M., and Evans, R. M. (2001) *Science* 294(5551), 2507-2511
18. Chevillard-Briet, M., Trouche, D., and Vandel, L. (2002) *EMBO Journal* 21(20), 5457-5466
19. Kwak, Y. T., Guo, J., Prajapati, S., Park, K. J., Surabhi, R. M., Miller, B., Gehrig, P., and Gaynor, R. B. (2003) *Molecular cell* 11(4), 1055-1066
20. Boulanger, M. C., Liang, C., Russell, R. S., Lin, R., Bedford, M. T., Wainberg, M. A., and Richard, S. (2005) *J Virol* 79(1), 124-131
21. Yu, M. C., Bachand, F., McBride, A. E., Komili, S., Casolari, J. M., and Silver, P. A. (2004) *Genes Dev* 18(16), 2024-2035
22. Bachand, F., and Silver, P. A. (2004) *Embo J* 23(13), 2641-2650
23. Fujiwara, T., Mori, Y., Chu, D. L., Koyama, Y., Miyata, S., Tanaka, H., Yachi, K., Kubo, T., Yoshikawa, H., and Tohyama, M. (2006) *Molecular and cellular biology* 26(6), 2273-2285
24. Li, H., Park, S., Kilburn, B., Jelinek, M. A., Henschen-Edman, A., Aswad, D. W., Stallcup, M. R., and Laird-Offringa, I. A. (2002) *J Biol Chem* 277(47), 44623-44630
25. Lukong, K. E., and Richard, S. (2004) *Nature structural & molecular biology* 11(10), 914-915
26. Smith, W. A., Schurter, B. T., Wong-Staal, F., and David, M. (2004) *J Biol Chem* 279(22), 22795-22798
27. Cote, J., Boisvert, F. M., Boulanger, M. C., Bedford, M. T., and Richard, S. (2003) *Mol Biol Cell* 14(1), 274-287
28. Nichols, R. C., Wang, X. W., Tang, J., Hamilton, B. J., High, F. A., Herschman, H. R., and Rigby, W. F. (2000) *Experimental cell research* 256(2), 522-532
29. Passos, D. O., Quaresma, A. J., and Kobarg, J. (2006) *Biochem Biophys Res Commun* 346(2), 517-525
30. Bedford, M. T., and Richard, S. (2005) *Molecular cell* 18(3), 263-272
31. Adams, M. M., Wang, B., Xia, Z., Morales, J. C., Lu, X., Donehower, L. A., Bochar, D. A., Elledge, S. J., and Carpenter, P. B. (2005) *Cell Cycle* 4(12), 1854-1861
32. Boisvert, F. M., Rhie, A., Richard, S., and Doherty, A. J. (2005) *Cell Cycle* 4(12), 1834-1841
33. El-Andaloussi, N., Valovka, T., Toueille, M., Steinacher, R., Focke, F., Gehrig, P., Covic, M., Hassa, P. O., Schar, P., Hubscher, U., and Hottiger, M. O. (2006) *Molecular cell* 22(1), 51-62
34. Bedford, M. T., Frankel, A., Yaffe, M. B., Clarke, S., Leder, P., and Richard, S. (2000) *J Biol Chem* 275(21), 16030-16036
35. Ostareck-Lederer, A., Ostareck, D. H., Rucknagel, K. P., Schierhorn, A., Moritz, B., Huttelmaier, S., Flach, N., Handoko, L., and Wahle, E. (2006) *J Biol Chem* 281(16), 11115-11125
36. Fackelmayer, F. O. (2005) *Trends in biochemical sciences* 30(12), 666-671
37. Leiper, J., Murray-Rust, J., McDonald, N., and Vallance, P. (2002) *Proceedings of the National Academy of Sciences of the United States of America* 99(21), 13527-13532

38. Wang, Y., Wysocka, J., Sayegh, J., Lee, Y. H., Perlin, J. R., Leonelli, L., Sonbuchner, L. S., McDonald, C. H., Cook, R. G., Dou, Y., Roeder, R. G., Clarke, S., Stallcup, M. R., Allis, C. D., and Coonrod, S. A. (2004) *Science* 306(5694), 279-283
39. Cuthbert, G. L., Daujat, S., Snowden, A. W., Erdjument-Bromage, H., Hagiwara, T., Yamada, M., Schneider, R., Gregory, P. D., Tempst, P., Bannister, A. J., and Kouzarides, T. (2004) *Cell* 118(5), 545-553
40. Lee, J., Cheng, D., and Bedford, M. T. (1990) *Techniques in Protein Methylation*, Humana Press Inc., Totowa, NJ
41. Pawlak, M. R., Banik-Maiti, S., Pietenpol, J. A., and Ruley, H. E. (2002) *J Cell Biochem* 87(4), 394-407
42. Chen, D. H., Wu, K. T., Hung, C. J., Hsieh, M., and Li, C. (2004) *J Biochem (Tokyo)* 136(3), 371-376
43. Lee, H. W., Kim, S., and Paik, W. K. (1977) *Biochemistry* 16(1), 78-85
44. Mowen, K. A., and David, M. (2001) *Sci STKE* 2001(93), PL1
45. Belyanskaya, L. L., Delattre, O., and Gehring, H. (2003) *Experimental cell research* 288(2), 374-381
46. Frankel, A., Yadav, N., Lee, J., Branscombe, T. L., Clarke, S., and Bedford, M. T. (2002) *J Biol Chem* 277(5), 3537-3543
47. Wada, K., Inoue, K., and Hagiwara, M. (2002) *Biochimica et biophysica acta* 1591(1-3), 1-10
48. Frankel, A., and Clarke, S. (2000) *Journal of Biological Chemistry* 275(42), 32974-32982
49. Lee, J. H., Cook, J. R., Yang, Z. H., Mirochnitchenko, O., Gunderson, S. I., Felix, A. M., Herth, N., Hoffmann, R., and Pestka, S. (2005) *J Biol Chem* 280(5), 3656-3664
50. Tang, J., Gary, J. D., Clarke, S., and Herschman, H. R. (1998) *Journal of Biological Chemistry* 273(27), 16935-16945
51. Ghosh, S. K., Paik, W. K., and Kim, S. (1988) *J Biol Chem* 263(35), 19024-19033
52. Lin, W. J., Gary, J. D., Yang, M. C., Clarke, S., and Herschman, H. R. (1996) *J Biol Chem* 271(25), 15034-15044
53. Strahl, B. D., Briggs, S. D., Brame, C. J., Caldwell, J. A., Koh, S. S., Ma, H., Cook, R. G., Shabanowitz, J., Hunt, D. F., Stallcup, M. R., and Allis, C. D. (2001) *Curr Biol* 11(12), 996-1000
54. Ma, H., Baumann, C. T., Li, H., Strahl, B. D., Rice, R., Jelinek, M. A., Aswad, D. W., Allis, C. D., Hager, G. L., and Stallcup, M. R. (2001) *Curr Biol* 11(24), 1981-1985
55. Boisvert, F. M., Cote, J., Boulanger, M. C., Cleroux, P., Bachand, F., Autexier, C., and Richard, S. (2002) *J Cell Biol* 159(6), 957-969
56. Komyod, W., Bauer, U. M., Heinrich, P. C., Haan, S., and Behrmann, I. (2005) *J Biol Chem* 280(23), 21700-21705
57. Belyanskaya, L. L., Gehrig, P. M., and Gehring, H. (2001) *Journal of Biological Chemistry* 276(22), 18681-18687
58. Zou, Y., and Wang, Y. (2005) *Biochemistry* 44(16), 6293-6301
59. Kim, S., Merrill, B. M., Rajpurohit, R., Kumar, A., Stone, K. L., Papov, V. V., Schneiders, J. M., Szer, W., Wilson, S. H., Paik, W. K., and Williams, K. R. (1997) *Biochemistry* 36(17), 5185-5192
60. Pahlich, S., Bschir, K., Chiavi, C., Belyanskaya, L., and Gehring, H. (2005) *Proteins* 61(1), 164-175
61. Rappsilber, J., Friesen, W. J., Paushkin, S., Dreyfuss, G., and Mann, M. (2003) *Anal Chem* 75(13), 3107-3114
62. Gehrig, P. M., Hunziker, P. E., Zahariev, S., and Pongor, S. (2004) *Journal of the American Society for Mass Spectrometry* 15(2), 142-149
63. Rouillon, A., Surdin-Kerjan, Y., and Thomas, D. (1999) *J Biol Chem* 274(40), 28096-28105
64. Xu, C., Henry, P. A., Setya, A., and Henry, M. F. (2003) *RNA (New York, N.Y* 9(6), 746-759
65. Najbauer, J., and Aswad, D. W. (1990) *J Biol Chem* 265(21), 12717-12721
66. Van Phi, L., and Soling, H. D. (1982) *Biochem J* 206(3), 481-487
67. Liu, Q., and Dreyfuss, G. (1995) *Molecular & Cellular Biology* 15(5), 2800-2808
68. Boisvert, F. M., Cote, J., Boulanger, M. C., and Richard, S. (2003) *Mol Cell Proteomics* 2(12), 1319-1330
69. Ong, S. E., Mittler, G., and Mann, M. (2004) *Nature methods* 1(2), 119-126
70. Khoury, J. D. (2005) *Advances in anatomic pathology* 12(4), 212-220
71. Delattre, O., Zucman, J., Plougastel, B., Desmaze, C., Melot, T., Peter, M., Kovar, H., Joubert, I., Dejong, P., Rouleau, G., Aurias, A., and Thomas, G. (1992) *Nature* 359(6391), 162-165
72. Zucman, J., Delattre, O., Desmaze, C., Epstein, A. L., Stenman, G., Speleman, F., Fletchers, C. D., Aurias, A., and Thomas, G. (1993) *Nature genetics* 4(4), 341-345
73. Ladanyi, M., and Gerald, W. (1994) *Cancer Res* 54(11), 2837-2840
74. Rauscher, F. J., 3rd, Benjamin, L. E., Fredericks, W. J., and Morris, J. F. (1994) *Cold Spring Harbor symposia on quantitative biology* 59, 137-146

75. Panagopoulos, I., Hoglund, M., Mertens, F., Mandahl, N., Mitelman, F., and Aman, P. (1996) *Oncogene* 12(3), 489-494
76. Alex, D., and Lee, K. A. (2005) *Nucleic Acids Res* 33(4), 1323-1331
77. Iakoucheva, L. M., Brown, C. J., Lawson, J. D., Obradovic, Z., and Dunker, A. K. (2002) *Journal of molecular biology* 323(3), 573-584
78. Deloulme, J. C., Prichard, L., Delattre, O., and Storm, D. R. (1997) *J Biol Chem* 272(43), 27369-27377
79. Zakaryan, R. P., and Gehring, H. (2006) *Journal of molecular biology* 363(1), 27-38
80. Lee, B. J., Cansizoglu, A. E., Suel, K. E., Louis, T. H., Zhang, Z., and Chook, Y. M. (2006) *Cell* 126(3), 543-558
81. Maris, C., Dominguez, C., and Allain, F. H. (2005) *The FEBS journal* 272(9), 2118-2131
82. Ohno, T., Ouchida, M., Lee, L., Gatalica, Z., Rao, V. N., and Reddy, E. S. (1994) *Oncogene* 9(10), 3087-3097
83. Lau, C. K., Diem, M. D., Dreyfuss, G., and Van Duyne, G. D. (2003) *Curr Biol* 13(11), 933-941
84. Zinszner, H., Sok, J., Immanuel, D., Yin, Y., and Ron, D. (1997) *J Cell Sci* 110 (Pt 15), 1741-1750
85. Iko, Y., Kodama, T. S., Kasai, N., Oyama, T., Morita, E. H., Muto, T., Okumura, M., Fujii, R., Takumi, T., Tate, S., and Morikawa, K. (2004) *J Biol Chem* 279(43), 44834-44840
86. Meyer, H. H., Wang, Y., and Warren, G. (2002) *Embo J* 21(21), 5645-5652
87. Ramos, A., Hollingworth, D., and Pastore, A. (2003) *RNA (New York, N.Y* 9(10), 1198-1207
88. Hovanessian, A. G., Puvion-Dutilleul, F., Nisole, S., Svab, J., Perret, E., Deng, J. S., and Krust, B. (2000) *Experimental cell research* 261(2), 312-328
89. Mikula, M., Dzwonek, A., Karczmarski, J., Rubel, T., Dadlez, M., Wyrwicz, L. S., Bomsztyk, K., and Ostrowski, J. (2006) *Proteomics* 6(8), 2395-2406
90. Olsen, R. J., and Hinrichs, S. H. (2001) *Oncogene* 20(14), 1756-1764
91. Ko, L. J., and Prives, C. (1996) *Genes Dev* 10(9), 1054-1072
92. Ayed, A., Mulder, F. A., Yi, G. S., Lu, Y., Kay, L. E., and Arrowsmith, C. H. (2001) *Nature structural biology* 8(9), 756-760
93. Xu, W., Doshi, A., Lei, M., Eck, M. J., and Harrison, S. C. (1999) *Molecular cell* 3(5), 629-638
94. Nakayama, J., Rice, J. C., Strahl, B. D., Allis, C. D., and Grewal, S. I. (2001) *Science* 292(5514), 110-113
95. Bertolotti, A., Melot, T., Acker, J., Vigneron, M., Delattre, O., and Tora, L. (1998) *Molecular and cellular biology* 18(3), 1489-1497
96. Rossow, K. L., and Janknecht, R. (2001) *Cancer Research* 61(6), 2690-2695
97. Thomas, G. R., Faulkes, D. J., Gascoyne, D., and Latchman, D. S. (2004) *Biochem Biophys Res Commun* 318(4), 1045-1051
98. Lee, J., Rhee, B. K., Bae, G. Y., Han, Y. M., and Kim, J. (2005) *Stem cells (Dayton, Ohio)* 23(6), 738-751
99. Knoop, L. L., and Baker, S. J. (2000) *J Biol Chem* 275(32), 24865-24871
100. Zhang, D., Paley, A. J., and Childs, G. (1998) *J Biol Chem* 273(29), 18086-18091
101. Yang, L., Embree, L. J., Tsai, S., and Hickstein, D. D. (1998) *J Biol Chem* 273(43), 27761-27764
102. Yang, L., Chansky, H. A., and Hickstein, D. D. (2000) *J Biol Chem* 275(48), 37612-37618
103. Chansky, H. A., Hu, M., Hickstein, D. D., and Yang, L. (2001) *Cancer Res* 61(9), 3586-3590
104. Zinszner, H., Albalat, R., and Ron, D. (1994) *Genes Dev* 8(21), 2513-2526
105. Rappsilber, J., Ryder, U., Lamond, A. I., and Mann, M. (2002) *Genome research* 12(8), 1231-1245
106. Kanai, Y., Dohmae, N., and Hirokawa, N. (2004) *Neuron* 43(4), 513-525
107. Angenstein, F., Evans, A. M., Ling, S. C., Settlage, R. E., Ficarro, S., Carrero-Martinez, F. A., Shabanowitz, J., Hunt, D. F., and Greenough, W. T. (2005) *J Biol Chem* 280(8), 6496-6503
108. Guipaud, O., Guillonneau, F., Labas, V., Praseuth, D., Rossier, J., Lopez, B., and Bertrand, P. (2006) *Proteomics* 6(22), 5962-5972
109. Hicks, G. G., Singh, N., Nashabi, A., Mai, S., Bozek, G., Klewes, L., Arapovic, D., White, E. K., Koury, M. J., Oltz, E. M., Van Kaer, L., and Ruley, H. E. (2000) *Nature genetics* 24(2), 175-179
110. Li, H., Watford, W., Li, C., Parmelee, A., Bryant, M. A., Deng, C., O'Shea, J., and Lee, S. B. (2007) *J Clin Invest*
111. Kuroda, M., Sok, J., Webb, L., Baechtold, H., Urano, F., Yin, Y., Chung, P., de Rooij, D. G., Akhmedov, A., Ashley, T., and Ron, D. (2000) *Embo J* 19(3), 453-462
112. Guinamard, R., Fougereau, M., and Seckinger, P. (1997) *Scandinavian journal of immunology* 45(6), 587-595
113. Felsch, J. S., Lane, W. S., and Peralta, E. G. (1999) *Curr Biol* 9(9), 485-488

2. Aim of the Thesis

Previous studies in our lab have revealed that the human RNA-binding protein EWS contains 30 dimethylated arginine residues situated in arginine-glycine rich regions (RGG boxes) within the RNA-binding domain (RBD). One goal of this study was to identify the protein arginine methyltransferases (PRMTs), which catalyze the post-translational methylation of the EWS protein. To date, nine human PRMTs have been described, which differ in substrate specificity and subcellular localization (nuclear, cytoplasmic, or cell membrane-associated). In vitro methylation experiments using recombinant EWS protein and selected recombinant PRMTs yield not only information about the enzyme that methylates the EWS protein but also first clues, where in the cell the methylation of the EWS protein takes place. The usual way of analyzing arginine methylation is the detection of incorporated radiolabeled methyl groups. In this study, mass spectrometry was chosen as main method of analysis. In comparison to radiolabeling, mass spectrometry yields not only the total amount of incorporated methyl groups, but also identifies the methylation sites, which is important for investigating sequence specificities of the different PRMTs.

The *EWS* gene undergoes tumor-related chromosomal translocations in the Ewing sarcoma family of tumors. The resulting oncogenic fusion proteins of the N-terminal EWS activation domain and C-terminal DNA-binding domains of ETS transcription factors have been investigated in detail. However, the function of the full-length proto-oncoprotein EWS in tumor suppression or in other conceivable cellular processes is largely unknown as the role of the extensively methylated RNA-binding domain of the EWS protein. To outline the spheres of action of the EWS protein, EWS-interacting proteins should be identified in the second part of this study. Large amounts of methylated and unmethylated EWS protein were a prerequisite for experiments on the co-purification, co-precipitation, and co-localization of potential EWS-binding proteins. As the yield of purified wild-type EWS protein from eukaryotic cell lines is rather poor, recombinant full-length EWS protein as well as diverse deletion mutants, all in unmethylated and methylated form, have been made available (in collaboration with Rouzanna Zakaryan). From the identification of the co-purified proteins by mass spectrometry, insights into the cellular roles of the multifunctional EWS protein and its post-translational modifications may be expected.

3. Different Methylation Characteristics of Protein Arginine Methyltransferase 1 and 3 towards the Ewing Sarcoma Protein and a peptide

Steffen Pahlich, Karim Bschir, Claudio Chiavi, Larisa Belyanskaya, and Heinz Gehring

Proteins. 2005 Oct 1;61(1):164-75.

Summary

The multifunctional Ewing Sarcoma (EWS) protein, a member of a large family of RNA-binding proteins, is extensively asymmetrically dimethylated at arginine residues within RGG consensus sequences. Using recombinant proteins we examined whether type I protein arginine methyltransferase (PRMT)1 or 3 is responsible for asymmetric dimethylations of the EWS protein.

After in vitro methylation of the EWS protein by GST-PRMT1, we identified 27 dimethylated arginine residues out of 30 potential methylation sites by mass spectrometry-based techniques (MALDI-TOF MS and MS/MS). Thus, PRMT1 recognizes most if not all methylation sites of the EWS protein. With GST-PRMT3, however, only 9 dimethylated arginines, located mainly in the C-terminal region of EWS protein, could be assigned, indicating that structural determinants prevent complete methylation. In contrary to previous reports this study also revealed that trypsin is able to cleave after methylated arginines. Pull-down experiments showed that endogenous EWS protein binds efficiently to GST-PRMT1 but less to GST-PRMT3, which is in accordance to the in vitro methylation results. Furthermore, methylation of a peptide containing different methylation sites revealed differences in the site selectivity as well as in the kinetic properties of GST-PRMT1 and GST-PRMT3. Kinetic differences due to an inhibition effect of the methylation inhibitor S-adenosyl-L-homocysteine could be excluded by determining the corresponding K_i values of the two enzymes and the K_d values for the methyl donor S-adenosyl-L-methionine.

The study demonstrates the strength of MS-based methods for a qualitative and quantitative analysis of enzymic arginine methylation, a posttranslational modification that becomes more and more the object of investigations.

Introduction

Methylation of arginine residues is catalyzed by protein arginine N-methyltransferases (PRMT), which use S-adenosyl-L-methionine (AdoMet) as donor of methyl groups [1]. Several cellular processes like subcellular localization of proteins, signal transduction, transcriptional regulation, and also the ability of proteins to bind to nucleic acids are affected by arginine methylation [2-7].

The type I PRMT catalyze the formation of N^G-monomethylarginine and asymmetric ω-N^G,N'^G-dimethylarginine, while type II enzymes form N^G-monomethylarginine and symmetric ω-N^G,N'^G-dimethylarginine [8]. Type I enzymes include the yeast methyltransferase Rmt1 [9], human PRMT1 [10], PRMT3 [11], coactivator-associated arginine methyltransferase (CARM1)/PRMT4 [12], and PRMT6 [13]. Several in vivo substrates for type I PRMT have been identified, among them fibrillarin, nucleolin, heterogeneous nuclear ribonucleoproteins (hnRNP), basic fibroblast growth factor, and histones [8]. PRMT5/JBP1 and PRMT7 are the only type II methyltransferases identified so far [14,15]. Substrates for symmetric arginine methylation known today are myelin basic proteins and Sm ribonucleoproteins [15-17].

The homologous type I arginine methyltransferases PRMT1 and PRMT3 used in this study differ not only in the N-terminal domain, where PRMT3 has a putative C_2H_2 type zinc-finger motif and a tyrosine phosphorylation consensus sequence, but also in substrate specificity, oligomerization, and subcellular localization [11]. Whereas PRMT1 is the main arginine methyltransferase in eukaryotic cells [18] with a wide substrate spectrum, only one physiological substrate of PRMT3, the 40S ribosomal protein S2, was identified so far [19]. Although PRMT1 and PRMT3 differ in their substrate specificity, they both dimethylated arginine residues within an RGG consensus sequence [11].

In cellular Ewing sarcoma (EWS) protein asymmetric dimethylation of arginines within such RGG domains was found [20]. This protein therefore is a potential substrate for the methyltransferases PRMT1 and PRMT3. The EWS protein consists of two main large domains, the N-terminal transcriptional activation domain (EAD) and the C-terminal RNA-binding domain (RBD). The EAD is involved in Ewing family tumors where chromosomal translocations lead to the formation of chimeric fusion products with oncogenic potential between the EAD coding region of the EWS gene and different ETS (erythroblastosis virus-transforming sequence) transcription factors [21-23]. However, only the N-terminal part of the EWS gene is involved in the translocation, and loss of RBD increases activation of EAD [24,25].

The function of native EWS protein remains unclear. The C-terminal RBD contains three arginine-glycine rich domains (RGG boxes), an RNA-recognition motif, and a C_2C_2 type zinc finger motif (see Fig. 3.4 C). It belongs to a large family of RNA-binding proteins such as hnRNP, mRNA poly(A)-binding proteins, and alternative splicing factors [26-29]. Although EWS protein is supposed to be a nuclear protein, it is also present in the cytosol and on the cell surface [20]. Independent on its localization, EWS protein isolated from cells is extensively dimethylated on arginine residues in the C-terminal RBD. Out of 30 arginines within possible methylation sites, 29 arginines were found to be asymmetrically dimethylated [20]. Recent studies indicated, that expression of EWS protein in the various subcellular compartments is affected by the methylation process, in particular on the availability of intracellular AdoMet [30]. However, the functional relevance of the extensive arginine methylation of the EWS protein as well as the methyltransferases, which are catalyzing the methylation of EWS protein are still unknown.

In this study, we demonstrate that the type I methyltransferases GST-PRMT1 and GST-PRMT3 have different methylation characteristics towards a synthetic peptide containing four RGG-methylation sites. In vitro methylation of the recombinant GST-EWS protein with GST-PRMT1 and GST-PRMT3 as well as pull-down experiments with cell extracts show that EWS protein interacts with both PRMT1 and PRMT3 but in a different manner. This underlines the differential substrate specificity of the two methyltransferases. MALDI-TOF MS and MS-based peptide sequencing (MS/MS) was used to monitor and identify specific arginine methylations.

Materials and Methods

Materials - The cDNA of EWS was a generous gift from Dr. Olivier Delattre (Institut Curie, Pathologie Moléculaire des Cancers, Paris Cedex). pGEX(SN)-PRMT1 and pGEX(SN)-PRMT3 constructs were kindly provided by Dr. Harvey R. Herschman (UCLA, Molecular Biology Institute, Los Angeles). E.coli lysate containing GST-PRMT6 was kindly obtained from Dr. Michael Hottiger (Institut für Veterinärbiochemie und Molekularbiologie, Universität Zürich).

Preparation, Expression and Purification of GST Fusion Proteins - The EWS cDNA [23] was PCR-amplified using primers 5'ATCTGGAATTCAAATGGCGTCCACGGATTACA-GTACCTATAGCCAAGCTGCAGC3' and 5'CTGGTAGTCAATGCAGCTCGAGGGGGT-

CTCTGCATCTAGTAGG3'. The resulting fragment was digested with EcoRI and XhoI and inserted into the GST fusion vector pGEX-6P-2 (Amersham Pharmacia Biotech). The GST fusion proteins were expressed and purified according to the manufacturer`s protocol (Amersham Pharmacia Biotech) using XL1-Blue competent cells (Stratagene) for GST-EWS and BL21 (DE3) competent cells (Novagen) for GST-PRMT1 and GST-PRMT3. The purification of each protein was checked by SDS-PAGE as shown for GST-EWS protein (Fig 3.3, inlet, molecular mass markers (kDa) are indicated). Both bands represent full-length GST-EWS protein, as determined by MS analysis. The reasons for differences in the electrophoretic mobility, however, are unclear. Purified proteins were stored at –80°C in 50 mM sodium phosphate, pH 7.4. The protein concentration was determined by UV-spectrometry, using the theoretically calculated ε_{280} (ExPASy), or by BioRad assay.

Activity Determination of GST-PRMT1 and GST-PRMT3 - To determine and compare the specific activities of GST-PRMT1 and GST-PRMT3, a synthetic peptide corresponding to residues 676–692 of human nucleolin (GRGGFGGRGGFRGGRGG-NH_2, ANAWA Biomedical Services & Products) was used as substrate in methylation assays. The peptide was incubated with GST-PRMT1 or GST-PRMT3 and AdoMet (Amersham) in a final volume of 50 µl in 10 mM sodium phosphate, pH 7.4 for various periods at 30°C (concentrations are denoted in the figure legends). The methylation reaction was stopped by addition of an equal volume of 50% (v/v) acetonitrile (ACN) / 0.1% (v/v) trifluoracetic acid (TFA). Chymotryptic digestion of the peptide was performed in 10 mM sodium phosphate, pH 7.9 with a final concentration of 1 µg chymotrypsin / ml for 15 min at 30°C.

Quantification of Methylation - Arginine methylation does not affect the net charge of the peptide. To test whether the ionisation efficiency of the methylated and non-methylated reference peptide during MALDI-TOF MS is the same, unmethylated and eight times methylated peptide was mixed in varying ratios. The measured ratios of the monoisotopic peak intensities in the MALDI-TOF spectrum correspond to the theoretic ratios. Therefore, peptide methylation can be quantified. The sum of monoisotopic peak intensities (unmethylated and methylated) in a spectrum was set equal to the starting peptide concentration. From the peak intensities in a spectrum the fraction of each methylation intermediate of the starting concentration was estimated. To calculate the total amount of incorporated methyl groups at a certain time point, the concentrations of all methylation intermediates at that time point were multiplied by their methylation state and added up.

Kinetic calculations **-** Kinetic values were calculated using hyperbolic curve fitting (SigmaPlot 8.0) of the measured values to the Michaelis-Menten equation. Competitive inhibition of GST-PRMT1 and GST-PRMT3 by AdoHcy was determined by measuring the initial rates of peptide methylation with GST-PRMT1 or GST-PRMT3 in the presence of AdoMet and varying concentrations of AdoHcy indicated in figure 3.1 E. The K_i values were calculated using the equation

$$K_i' = K_i\left(1+\frac{[S]}{K_d}\right)$$

where the apparent K_i' corresponds to the inhibitor concentration at 50% inhibition (Fig. 3.1 E). K_i is the inhibition constant, [S] denotes the concentration of AdoMet and K_d the dissociation constant of AdoMet. K_d values of GST-PRMT1 and GST-PRMT3 for AdoMet were determined by measuring the decrease of tryptophan fluorescence at varying concentrations of AdoMet. Excitation was set at 285 nm and emission spectra from 295 to 420 nm were recorded. The fluorescence change at λ_{max} (343 nm) was plotted against the concentration of AdoMet and the K_d values were estimated by hyperbolic curve fitting.

Methylation of Recombinant GST-EWS Protein **-** GST-EWS was incubated with GST-PRMT1 or GST-PRMT3 and AdoMet in a final volume of 50 µl 10 mM sodium phosphate, pH 7.4 for 3 h at 30 °C (concentrations are denoted in the figure legends). Addition of 10 µl of two fold Laemmli sample buffer to 10 µl of the reaction mixture and heating for 5 min at 95°C terminated the reaction.

In-gel Digestion and Mass Spectrometry **-** Gel pieces containing the coomassie-stained protein band of GST-EWS were excised, destained twice in 100 µl 40% ACN / 0.1% TFA in 100 mM ammoniumbicarbonate for 15 min, washed in 100 µl water for 15 min, dehydrated in 100% ACN for 5 min and dried completely. In-gel digestion was carried out with trypsin (Promega, 6 µg/ml) or chymotrypsin (Promega, 5 µg/ml) overnight at room temperature in 15 µl 0.5 mM Tris-HCl, pH 8.0. One µl of the digest was mixed with 3 µl of matrix solution (α-cyano-4-hydroxycinnamic acid (α-CCA) saturated in 40% ACN / 0.1% TFA) and spotted directly onto a MALDI target. MALDI-TOF MS was performed at the Protein Analysis Unit (Universität Zürich) on a Bruker Biflex instrument in the reflector mode employing pulsed ion extraction. The spectra were mass calibrated externally. To assign and identify methylated peptides, spectra of native EWS protein [20] but also theoretic calculated digests of EWS protein (PeptideMass from ExPASy) were used as reference. The methylated peptides were assigned

according to differences (addition of 14 Da per methylation) observed in spectra of PRMT treated and untreated EWS protein.

Reverse-Phase HPLC and MALDI-TOF MS/MS **-** RP-HPLC was performed at the Functional Genomics Center Zurich (FGCZ) on an LC-Packings system as described elsewhere [31] with some modifications. The sample was desalted using C_{18} ZipTips (Millipore, Bedford, MA), peptides were eluted in 10 µl of 50% ACN / 0.1% TFA, dried completely and resolved in 10 µl of solvent A (1% ACN / 0.1% TFA). The sample was loaded directly to the system and peptides were separated with a flow rate of 0.3 µl/min using the following gradient: 0-5 min, 0% solvent B (80% ACN / 0.1% TFA); 5-35 min, 0-50% B; 35-40 min, 50-100% B, followed by 100% B. The HPLC eluent was mixed with MALDI matrix (3 mg α-CCA in 70% ACN / 0.1% TFA) at a flow rate of 1 µl/min before being spotted every 20 s on a blank MALDI plate (Applied Biosystems). LC-samples, spotted on a MALDI target, were analyzed at the FGCZ on a 4700 Proteomics Analyzer MALDI-TOF/TOF (Applied Biosystems) as described [31] using 1500 shots per MS spectrum and 3000-6000 shots per MS/MS spectrum. Maximal 10 MS/MS spectra were recorded per sample spot from precursor ions with a minimal signal-to-noise (S/N) ratio of 40 in the mass spectrum. GPS Explorer (Applied Biosystems) was used for submitting the data acquired from TOF/TOF for database searching using Mascot (www.matrixscience.com). The peptide tolerance was set to 75 ppm and the MS/MS tolerance to 0.2 Da.

GST Pull-down Experiments **-** Purified GST-PRMT1 or GST-PRMT3 (10 µg each) was incubated for 1 h at 4°C with 30 µl of glutathione-agarose (Sigma). Cell extracts from Jurkat Cells and H9 cells were fractionated as described before [30]. Nuclear cell extracts (150 µg protein in 450 µl EBC buffer: 50mM Tris-HCl, 120 mM sodium chloride, 0.5% (v/v) Nonidet P-40, 1mM 1,4-Dithio-DL-threitol, pH 8.0) were incubated for 1 h at 4°C with agarose-bound GST-PRMT1, GST-PRMT3, or GST alone. The agarose was washed three times with EBC buffer and bound proteins were extracted by incubation with two fold Laemmli sample buffer for 5 min at 95°C.

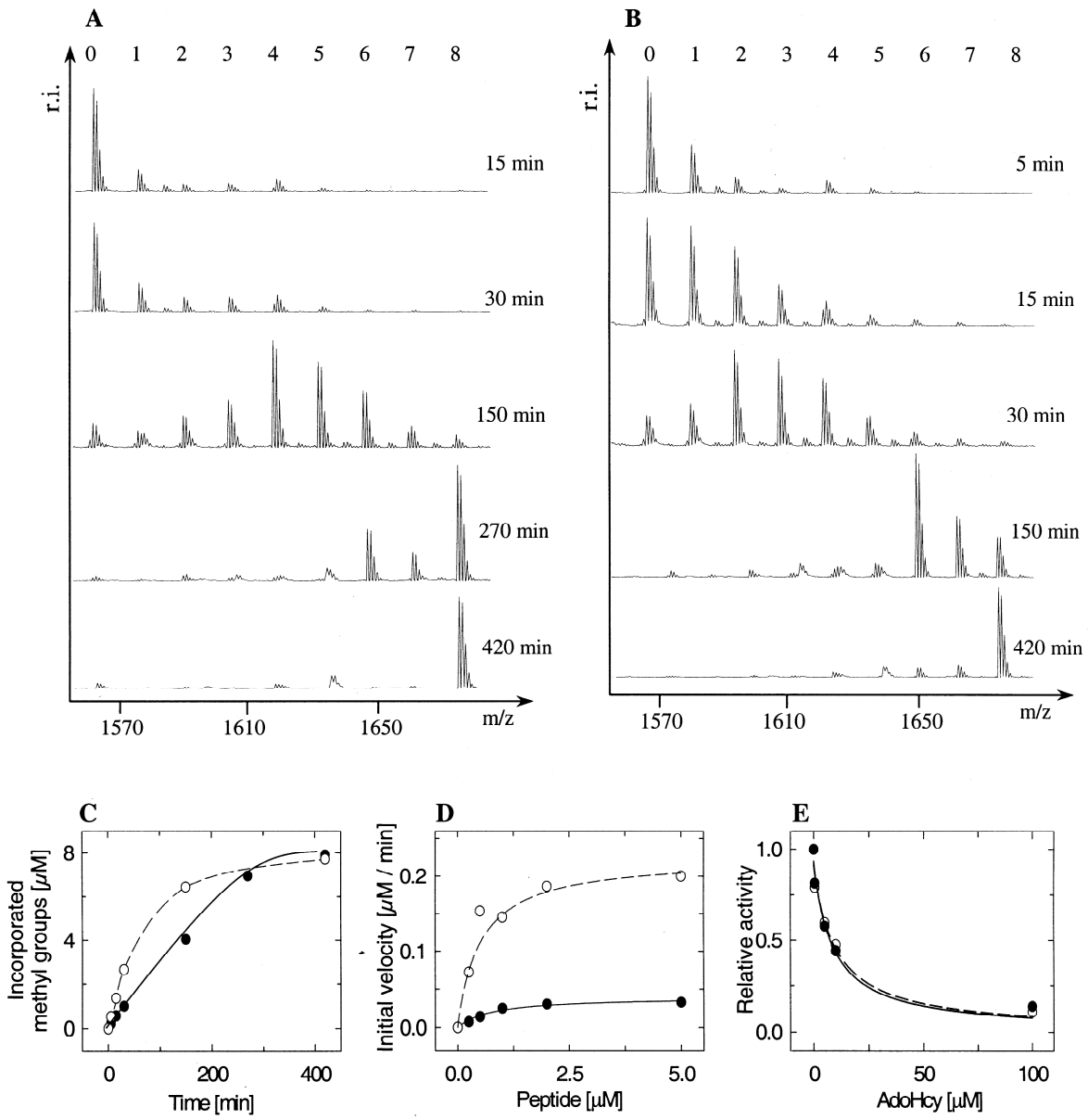


Figure 3.1: Time course of methylation of the peptide by GST-PRMT1 **(A)** and GST-PRMT3 **(B)**. Peptide (1 μM) was incubated with 0.1 μM of each enzyme and AdoMet (100 μM) and analyzed by MALDI-TOF MS at the indicated time points. The theoretical m/z value of the peptide is 1563.72 (unmethylated) or 1675.93 (completely methylated). All other peaks correspond to methylation intermediates (+14 per incorporated methyl group) and the numbers indicate the amount of incorporated methyl groups. **(C)**: Kinetics of incorporation of the eight methyl groups per peptide by GST-PRMT1 (●) or GST-PRMT3 (○). **(D)**: Dependence of the initial rates of peptide methylation with 0.1 μM GST-PRMT1 (●) or 0.1 μM GST-PRMT3 (○) on the concentration of the peptide. **(E)**: Inhibition of methyltransferase activity by AdoHcy. Methylation was performed with 0.1 μM GST-PRMT1 (●) or 0.1 μM GST-PRMT 3 (○), 1 μM peptide, and 10 μM AdoMet in the absence and presence of AdoHcy. Initial velocities at different concentrations of AdoHcy were measured and the relative activities are indicated.

Results

Comparison of Activities of GST-PRMT1 and GST-PRMT3 in Peptide Methylation

To prove and compare the activities of GST-PRMT1 and GST-PRMT3, a synthetic peptide (GRGGFGGRGGFRGGRGG-NH_2) containing four unmethylated arginine residues within an RGG sequence was used as substrate for in vitro methylation (Materials and Methods). The methylation of the peptide by GST-PRMT1 and GST-PRMT3 was monitored at different time points by MALDI-TOF MS. Both enzymes showed activity towards the peptide but with distinct differences in the pattern of methylation during the time course of the reaction.

The incorporation of methyl groups at all four arginines by GST-PRMT1 proceeded linearly until the peptide was methylated almost completely (Fig. 3.1 A and C). By plotting initial rates of methylation of the unmethylated peptide at different peptide concentrations (Fig. 3.1 D) kinetic constants were calculated resulting in a k_{cat}/K_m value of $8.1*10^{-3}$ $\mu M^{-1}s^{-1}$ (Table I, for calculations see Materials and Methods). Peptide methylation with GST-PRMT3 was initially faster than with GST-PRMT1 (Fig. 3.1 B and C), already after 15 minutes, all eight methylation intermediates could be observed. After an initial linear phase, the rate decreased steadily until complete methylation of the peptide was attained after ~7 hours. The initial rates of peptide methylation by GST-PRMT3 at different substrate concentrations yielded a k_{cat}/K_m value of 0.1 $\mu M^{-1}s^{-1}$ which is ~12 times higher than the value of GST-PRMT1 (Fig. 3.1 D; Table I).

The origin of the different peptide methylation characteristics could be due to a differential inhibition effect of S-adenosyl-L-homocysteine (AdoHcy), the product of the methyl donor AdoMet. AdoHcy is known to be a competitive inhibitor of arginine methyltransferases [1]. In a cellular environment, hydrolysis of AdoHcy to adenosine and homocysteine by AdoHcy-hydrolase, which was not present in the in vitro methylation reactions, prevents accumulation of AdoHcy.

In order to determine whether inhibition by AdoHcy arising from the peptide methylation influences the activity of the PRMT, peptide methylation was performed in the absence and presence of varying concentrations of AdoHcy (Fig. 3.1 E). Increasing amounts of AdoHcy inhibited the methylation activity of both GST-PRMT1 and GST-PRMT3 under the present conditions with an apparent K_i' of 8.6 μM and 9.7 μM, respectively. For calculating the competitive inhibition constant K_i, the K_d values of the PRMT for AdoMet were determined by measuring the quenching of tryptophan fluorescence of PRMT on addition of AdoMet (Materials and Methods). The K_d values of GST-PRMT1 and GST-PRMT3 for AdoMet and the calculated competitive inhibition constants K_i (Table I) indicated similar affinities of each methyltransferase for the substrate AdoMet and for the product AdoHcy. Under the conditions used in Fig. 3.1 A-C (100 μM AdoMet), the initial rates of peptide methylation are hardly influenced by the formation of AdoHcy. Even after complete methylation of 1 μM of peptide, AdoMet remains in a great excess compared to AdoHcy resulting in a negligible inhibition.

Thus, other reasons must be responsible for the decline of methylation rate in the case of GST-PRMT3. Three of the four arginines of the peptide are embedded in a GRGG sequence motif whereas one arginine has a phenylalanine at its N-terminal site (FRGG). In order to investigate whether GST-PRMT3 has any sequence preferences or whether the arginine residues of the peptide are methylated randomly, the methylation reaction was stopped before completion (Fig. 3.2 A, left). At that time point, mainly six-, seven-, and eight-fold methylated peptide was found. Chymotryptic digestion of the peptide yielded the peptide fragments 1-11 and 12-17 with masses of 1024.5 Da and 558.7 Da, respectively, in the unmethylated state. Only a four-fold methylated state could be observed in fragment 1-11, whereas fragment 12-17 is present with three peaks representing the two-, three-, and four-fold methylated state (Fig. 3.2 A, right). Obviously, arginines 2 and 8 were both dimethylated at that time point, whereas one of the arginines 12 and 15 was dimethylated and one was present in the un-, mono-, and dimethylated state. Most likely it is Arg 12 (FRGG) that differs in

TABLE I. Kinetic properties of GST-PRMT1 and GST-PRMT3 towards the peptide, AdoMet, and AdoHcy.

	GST-PRMT1	GST-PRMT3
V_{max}	0.04 μM / min	0.22 μM / min
K_m	0.8 μM	0.35 μM
k_{cat}	$6.5 * 10^{-3}$ s^{-1}	$3.6 * 10^{-2}$ s^{-1}
k_{cat} / K_m	$8.1 * 10^{-3}$ $\mu M^{-1}s^{-1}$	0.1 $\mu M^{-1}s^{-1}$
K_d (AdoMet)	1.1 μM	3.9 μM
K_i (AdoHcy)	0.8 μM	2.7 μM

Values were calculated as described (Materials and Methods) using hyperbolic curve fitting of Fig. 3.1 D-E.

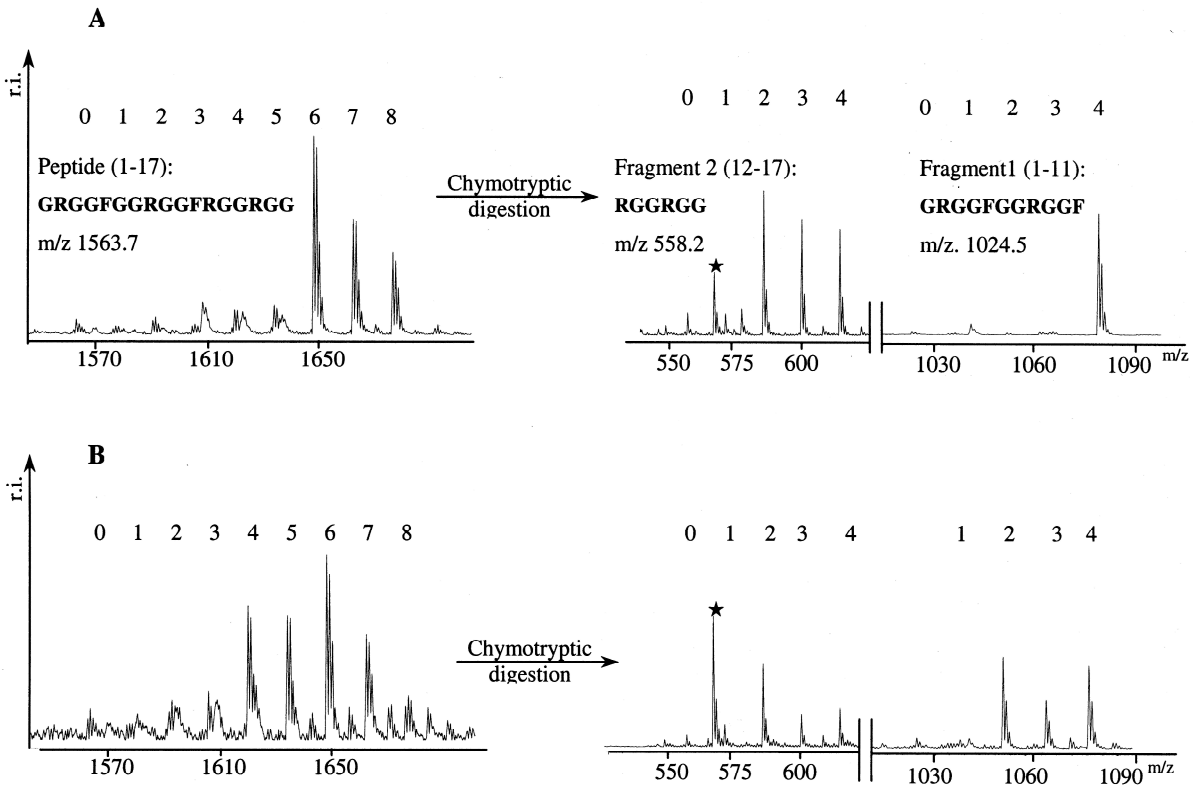


Figure 3.2: Analysis of sequential methylation of the peptide. Methylation of the peptide with GST-PRMT3 **(A)** or GST-PRMT1 **(B)** was stopped before completion and analyzed with MALDI-TOF MS (left). The peptide was cleaved with chymotrypsin and the resulting fragments (1-11; 12-17) were subjected to MS analysis to estimate their methylation state (right). The numbers indicate the incorporated methyl groups. The peak at m/z 568 labeled with an asterisk is a signal arising from the matrix (α-CCA).

sequence from the others. Thus, arginines within a GRGG sequence (Arg 2, 8, and 15) motif are probably methylated faster by GST-PRMT3 than those in a FRGG (Arg 12) motif. The same experiment was performed with GST-PRMT1 (Fig. 3.2 B). The mass spectrum of the chymotryptic fragments of incomplete methylated peptide (Fig. 3.2 B, right) showed in both fragments the same methylation distribution of two, three, and four methyl groups per peptide fragment, indicating that GST-PRMT1 methylates in a random order without sequence preference.

Methylation of GST-EWS Protein

Given the activity of GST-PRMT1 and GST-PRMT3 towards the reference peptide, we investigated whether GST-EWS protein, which is extensively methylated in vivo [20], is a substrate for the two methyltransferases. Recombinant GST-EWS protein was incubated with GST-PRMT1 in the presence of AdoMet, the reaction was stopped after 3 h and GST-EWS protein was separated from GST-PRMT1 by SDS-PAGE. The band representing the full-

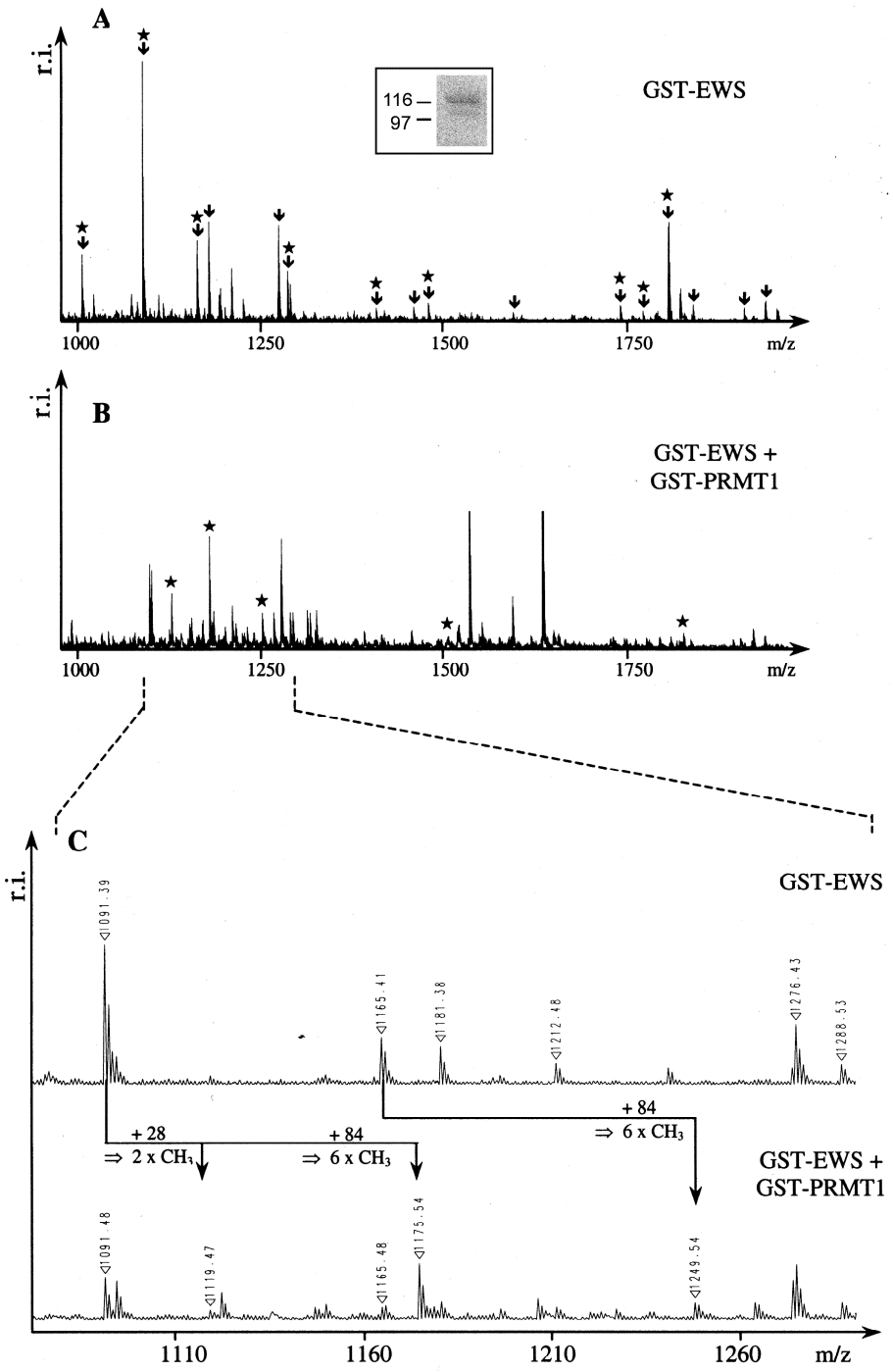


Figure 3.3: MALDI-TOF mass spectra of in vitro methylated GST-EWS by GST-PRMT1. GST-EWS (2 μM) was incubated with GST-PRMT1 (1 μM) and AdoMet (1 mM) for 3 h at 30°C and subjected to SDS-PAGE. The dominant band after Coomassie blue staining representing the full-length GST-EWS protein (inlet) was cut out and in-gel digested with chymotrypsin. **(A)** MALDI-TOF mass spectrum of unmodified GST-EWS protein. Peaks labeled with an arrow could be assigned to GST-EWS protein; peaks additionally labeled with an asterisk are EWS peptides containing RGG consensus sequences. **(B)** MALDI-TOF mass spectrum of GST-EWS protein methylated by GST-PRMT1. Peaks with a mass shift corresponding to the addition of methyl groups are indicated with an asterisk. **(C)** Enlarged spectra of A and B.

length GST-EWS protein was cut out and subjected to tryptic or chymotryptic in-gel digestion. The resulting peptides were analyzed by MALDI-TOF MS as described [20]. In the MALDI spectrum of untreated GST-EWS protein, all dominant peaks could be assigned to the GST-EWS protein (Fig. 3.3 A). In the mass fingerprint of GST-EWS protein treated with GST-PRMT1 several mass shifts could be observed (Fig. 3.3 B). Dominant peaks of the untreated GST-EWS protein decreased after GST-PRMT1 treatment and new peaks corresponding to the addition of methyl groups appeared (Fig. 3.3 C).

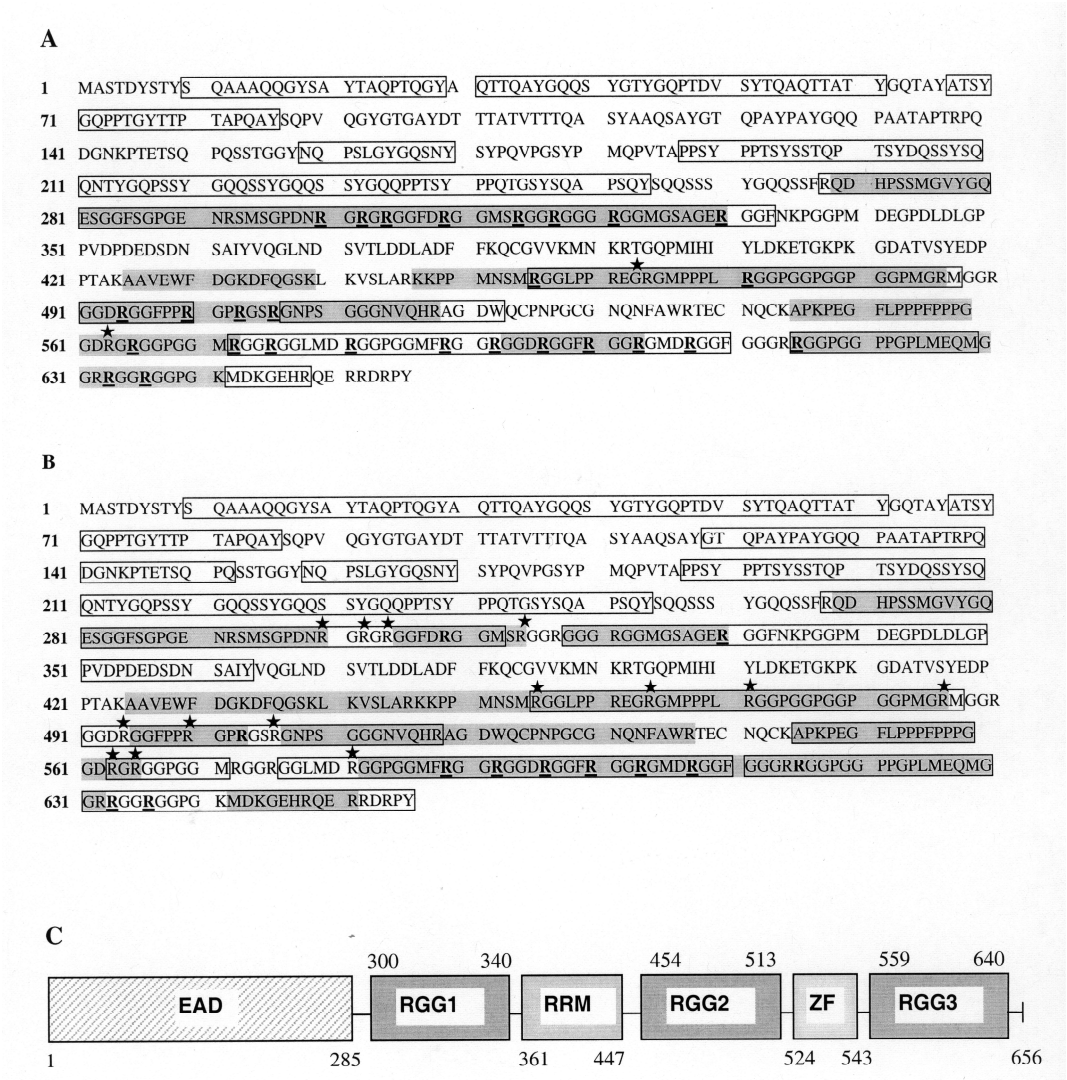


Figure 3.4: Sequence coverage by identified peptides of the EWS protein methylated either with GST-PRMT1 **(A)** or GST-PRMT3 **(B)** for 3 hours. The shaded regions represent tryptic peptides, chymotryptic peptides are labeled with a box. Monometylated arginines are bold (R), arginines found to be dimethylated are bold and underlined (R). Identified unmodified methylation sites are labeled with an asterisk («). **(C)**: Domains of the EWS protein and the corresponding numbers of amino acid residues. EAD – EWS activation domain; RGG1-3 – arginine-glycine rich domains 1-3; RRM – RNA recognition motif; ZF – putative zinc finger domain.

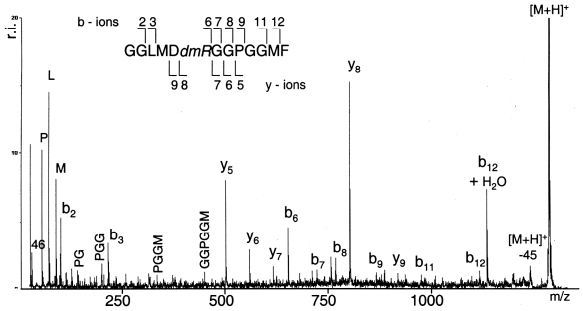


Figure 3.5: MS/MS spectrum of the chymotryptic peptide 576 – 588 (m/z 1279.6) of the EWS protein methylated with PRMT1. The labeled y- and b-fragment ions as well as the internal fragments cover the sequence including the dimethylated arginine (dmR) 581. Ions at m/z 46 and at m/z 1234.6 ($[M+H]^+$-45) resulting from the loss of dimethylammonium are additional proofs for the asymmetric dimethylation of the arginine residue [32,33].

Chymotryptic digestion is not as specific as tryptic digestion, resulting in a larger number of possible EWS fragments. Additional modifications like methionine oxidation (mass increase of 16 Da), arginine monomethylation (+14 Da), or dimethylation (+28 Da) even increase the number of theoretically generated peptides, among them several peptides with identical masses. To confirm the identity of the matched tryptic and chymotryptic peptides, the digests were subjected to RP-HPLC followed by MALDI-TOF MS/MS sequencing of the peptides (see Materials and Methods). Taken together, the analyzed tryptic and chymotryptic peptides covered a large part of the EWS sequence (Fig. 3.4 A), and 27 dimethylated arginines were found and confirmed by MS/MS analysis after GST-PRMT1 treatment (Table II). In MS/MS spectra, as exemplified on the chymptryptic peptide 576 – 588 including a dimethylated arginine, all dominant peaks could be assigned as b- or y-fragment ions or internal fragments (Fig. 3.5). The appearance of specific fragment ions resulting from the loss of dimethylammonium (m/z 46) and the release of dimethylammonium from the precursor ion ($[M+H]^+$-45) confirmed the asymmetric dimethylation of arginine residues [32,33]. Arginines 464 and 563 were found to be unmodified and arginine 490 was not covered by the identified peptides (Fig. 3.4). Some peptides were present in different methylation states (e.g. chymotryptic peptide 589-599) indicating, that the methylation reaction was not complete after 3 h. Nevertheless, almost all methylation sites were recognized by GST-PRMT1.

Although GST-PRMT3 has a higher activity than GST-PRMT1 towards the reference peptide, no methylation activity towards the EWS protein was found when comparable activity units of GST-PRMT3 were used as it was done with GST-PRMT1. After treatment of the GST-EWS protein (2 μM) with 1 μM of GST-PRMT3, mass shifts could be observed as well, but much fewer dimethylated arginines were found. Again, a large part of the EWS sequence was covered by the identified tryptic and chymotryptic peptides (Fig. 3.4 B), but only nine dimethylated and three monomethylated arginines were found by MS/MS (Table

Table II: Tryptic and chymotryptic peptides of GST-EWS, methylated with GST-PRMT1, identified by MS/MS. Only methyl-arginine containing pepides are listed. Monomethylated arginines are indicated in bold (**R**), dimethylated arginines are bold and underlined (**R**). The presence of the fragment ion at m/z 46 indicates asymmetric dimethylation.

tryptic peptides					chymotryptic peptides				
predicted mass	observed mass	residue numbers	Sequence	m/z 46	predicted mass	observed mass	residue numbers	Sequence	m/z 46
1090.51	1090.52	293 to 302	SMSGPDN**R**GR		1787.82	1787.93	286 to 302	SGPGENRSMSGPDN**R**GR	
1104.52	1104.53	293 to 302	SMSGPDN**R**GR	ü	1801.84	1801.96	286 to 302	SGPGENRSMSGPDN**R**GR	ü
1132.55	1132.58	293 to 302	SMSGPDN**R**G**R**	ü	1132.55	1132.60	293 to 302	SMSGPDN**R**G**R**	ü
1373.71	1373.74	293 to 304	SMSGPDN**R**G**R**G**R**	ü	1373.71	1373.79	293 to 304	SMSGPDN**R**G**R**G**R**	ü
1634.87	1634.93	303 to 317	G**R**GGFD**R**GGMS**R**GG**R**	ü	1023.48	1023.54	303 to 312	G**R**GGFDRGGM	
1067.51	1067.53	303 to 317	GGFD**R**GGMSR	ü	1037.49	1037.56	303 to 312	G**R**GGFDRGGM	ü
1393.71	1393.76	305 to 317	GGFD**R**GGMS**R**GG**R**	ü	1065.53	1065.58	303 to 312	G**R**GGFD**R**GGM	ü
1474.73	1474.77	315 to 330	GG**R**GGG**R**GGMGSAGER	ü	824.37	824.41	305 to 312	GGFD**R**GGM	ü
1176.55	1176.57	318 to 330	GGG**R**GGMGSAGER	ü	1393.71	1393.79	305 to 317	GGFD**R**GGMS**R**GG**R**	ü
988.51	988.53	448 to 455	KPPMNSM**R**	ü	1487.80	1487.89	308 to 321	D**R**GGMS**R**GG**R**GGG**R**	ü
1984.00	1984.11	465 to 486	GMPPPL**R**GGPGGPGGPGGPMGF	ü	1437.67	1437.76	318 to 333	GGG**R**GGMGSAGERGGF	ü
1381.68	1381.59	491 to 503	GGD**R**GGFPPRGP**R**	ü	1096.48	1096.54	322 to 333	GGMGSAGE**R**GGF	
1737.96	1738.03	491 to 506	GGD**R**GGFPP**R**GP**R**GS**R**	ü	1110.50	1110.56	322 to 333	GGMGSAGE**R**GGF	ü
1845.95	1846.00	501 to 518	GP**R**GS**R**GNPSGGGNVQHR	ü	1122.65	1122.71	455 to 464	**R**GGLPPREGR	ü
1507.75	1507.78	504 to 518	GS**R**GNPSGGGNVQHR	ü	1756.86	1756.98	465 to 484	GMPPPL**R**GGPGGPGGPGGPM	
872.42	872.39	564 to 572	G**R**GGPGGMR	ü	1770.88	1771.00	465 to 484	GMPPPL**R**GGPGGPGGPGGPM	ü
849.45	849.49	593 to 600	GGD**R**GGFR	ü	1178.57	1178.65	471 to 484	**R**GGPGGPGGPGGPM	ü
1175.64	1175.67	593 to 603	GGD**R**GGF**R**GG**R**	ü	1391.70	1391.77	471 to 486	**R**GGPGGPGGPGGPMGR	ü
1778.88	1778.99	615 to 632	**R**GGPGGPPGPLMEQMGGR	ü	1043.54	1043.59	491 to 500	GGD**R**GGFPPR	ü
897.47	897.50	633 to 641	**R**GG**R**GGPGK	ü	1761.90	1762.01	572 to 588	**R**GG**R**GGLMD**R**GGPGGMF	ü
					1265.58	1265.64	576 to 588	GGLMD**R**GGPGGMF	
					1279.59	1279.66	576 to 588	GGLMD**R**GGPGGMF	ü
					921.42	921.47	580 to 588	D**R**GGPGGMF	ü
					1119.57	1119.63	589 to 599	RGG**R**GGDRGGF	ü
					1147.61	1147.66	589 to 599	**R**GG**R**GGDRGGF	ü
					1175.64	1175.70	589 to 599	**R**GG**R**GGD**R**GGF	ü
					1249.66	1249.72	600 to 610	**R**GG**R**GMD**R**GGF	ü
					1120.59	1120.65	615 to 626	**R**GGPGGPPGPLM	ü
					1508.73	1508.85	615 to 629	**R**GGPGGPPGPLMEQM	ü

Table III: tryptic and chymotryptic peptides of GST-EWS, methylated with GST-PRMT3, identified by MS/MS. Only methyl-arginine containing pepides are listed. Monomethylated arginines are indicated in bold (**R**), dimethylated arginines are bold and underlined (**R**). The presence of the fragment ion at m/z 46 indicates asymmetric dimethylation.

tryptic peptides					chymotryptic peptides				
predicted mass	observed mass	residue numbers	Sequence	m/z 46	predicted mass	observed mass	residue numbers	Sequence	m/z 46
1053.49	1053.54	305 to 314	GGFD**R**GGMSR		1023.48	1023.52	303 to 312	GRGGFD**R**GGM	
1162.54	1162.59	318 to 330	GGGRGGMGSAGE**R**		1423.65	1423.73	318 to 333	GGGRGGMGSAGE**R**GGF	
1062.53	1062.60	582 to 592	GGPGGMF**R**GGR		1437.67	1437.75	318 to 333	GGGRGGMGSAGE**R**GGF	ü
1076.54	1076.59	582 to 592	GGPGGMF**R**GGR	ü	1096.48	1096.54	322 to 333	GGMGSAGE**R**GGF	
1090.56	1090.61	582 to 592	GGPGGMF**R**GG**R**	ü	1110.50	1110.54	322 to 333	GGMGSAGE**R**GGF	ü
1104.57	1104.61	582 to 592	GGPGGMF**R**GG**R**	ü	2065.04	2065.02	496 to 516	GFPPRGP**R**GSRGNPSGGGNVQ	
849.43	849.46	593 to 600	GGD**R**GGFR	ü	1105.56	1105.60	589 to 599	RGG**R**GGDRGGF	
1119.58	1119.62	593 to 603	GGD**R**GGFRGGR	ü	1119.58	1119.61	589 to 599	RGG**R**GGDRGGF	ü
1133.59	1133.64	593 to 603	GGD**R**GGF**R**GGR	ü	1147.61	1147.65	589 to 599	RGG**R**GGD**R**GGF	ü
1147.61	1147.65	593 to 603	GGD**R**GGF**R**GGR	ü	1179.58	1179.61	600 to 610	RGG**R**GMDRGGF	
1080.50	1080.54	604 to 614	GMD**R**GGFGGGR		1221.63	1221.68	600 to 610	**R**GG**R**GMDRGGF	ü
1094.52	1094.56	604 to 614	GMD**R**GGFGGGR	ü	1167.68	1167.69	630 to 641	GGR**R**GG**R**GGPGK	ü
1764.86	1764.97	615 to 632	**R**GGPGGPPGPLMEQMGGR						

III). Eight dimethylated and one monomethylated arginine were found in the C-terminal RGG box 3 (Fig 3.4 C), whereas the RGG boxes 1 and 2 of EWS protein were almost unmodified. Arginines 317, 490, 572, and 575 were not covered by the identified peptides.

Remarkably, we identified several tryptic peptides with dimethylated arginines at their C-termini although it was reported that trypsin does not cleave after a dimethylated arginine [16,34].

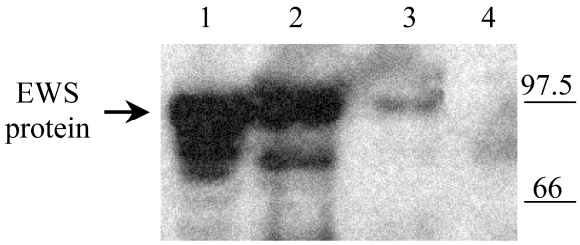


Figure 3.6: GST-pull down experiment using GST-PRMT1 and GST-PRMT3 as bait. Nuclear cell extracts from Jurkat cells (Materials and Methods) were used for pull-down experiments. Extracted proteins were subjected to SDS-PAGE and Western blot analysis using anti-EWS antibodies [30]. EWS protein from a nuclear cell extract (lane 1) was extracted with GST-PRMT1 (lane 2), with GST-PRMT3 (lane 3), or with GST alone (lane 4). Molecular mass markers are indicated.

Thus, arginine dimethylation might reduce, but not prevent cleavage by trypsin and the tryptic cleavage does not allow the conclusion that C-terminal arginine residues are unmethylated.

GST-pull down experiments

To investigate whether the in vitro methylation of EWS protein by PRMT1 and 3 reflects intracellular interactions, GST-pull down experiments were performed using GST-PRMT1 or GST-PRMT3 as baits. Jurkat cells were fractionated as described [30] and GST-PRMT1, GST-PRMT3, or GST pre-bound to glutathione (GSH)-agarose was added to the nuclear extract, which contains the highest amount of EWS protein. Proteins extracted with the GSH-agarose were subjected to SDS-PAGE, followed by Western blot analysis as described [30]. With GST-PRMT1 as bait, a large amount of EWS protein was extracted whereas with GST-PRMT3, only a weak EWS protein signal was present. With GST alone, no signal of EWS protein was detected (Fig. 3.6). Comparable results were obtained using the membrane and the cytosolic fraction of Jurkat cells or H9 cells (data not shown).

Discussion

In this study we investigated the methylation characteristics of the type I methyltransferases PRMT1 and PRMT3 towards a reference peptide with four methylation sites and towards the EWS protein containing 30 potential methylation sites. For both enzymes, the methylation reactions were monitored by MALDI-TOF MS. Using MS-based methods for analysing protein arginine methylation has the advantage in comparison to measuring incorporation of radiolabeled methyl groups, that not only the sites of methylation but also the degree of methylation (un-, mono-, or dimethylated) can be followed in a time-dependent manner. Methylation of the synthetic reference peptide proved the activity of both recombinant

enzymes GST-PRMT1 and GST-PRMT3 and revealed certain information on their mechanisms. GST-PRMT1 adds methyl groups to the peptide with a relatively low catalytic efficiency (Table I), values not known so far. The linear time course (Fig. 3.1 C) as well as the MS analysis of methylation intermediates by peptide fragmentation (Fig. 3.2 B) demonstrate that methylation by GST-PRMT1 takes place in a random order, independent of the sequence.

The methylation characteristics of GST-PRMT3 are different. The catalytic efficiency of GST-PRMT3 towards GRGG methylation sites is ~12 times higher than that of GST-PRMT1 (Table I). The reaction velocity decreases, however, on methylation of FRGG sequence motifs as revealed by MALDI-TOF MS of partly methylated peptide. Apparently, GST-PRMT3 differentiates between sequence motifs and methylates arginines within a GRGG motif most likely randomly and faster than those in a FRGG motif. Differences due to competitive inhibition by AdoHcy, the product of AdoMet, were excluded by inhibition studies. AdoHcy inhibits GST-PRMT1 and GST-PRMT3 in a comparable manner (Table I). The kinetic values presented here are in the same range as data for the protein methylase I purified from calf brain (K_d of 5.9 μM for AdoMet; K_i of 2.6 μM for AdoHcy) [1]. The kind of PRMT used in that study, however, could not be distinguished at that time.

The present results also demonstrate that GST-EWS protein is efficiently methylated by GST-PRMT1. Previous results have shown that at least 29 out of 30 arginines within potential methylation sites are dimethylated in vivo [20]. No information about the methylation state of the 30th arginine (Arg 330) was available, because it could not be detected in that study. Here we identified 27 dimethylated arginines after in vitro methylation by GST-PRMT1, among them Arg 330. Although one arginine could not be detected and two arginines were found to be unmodified, indicating that the methylation reaction was not complete after 3 h, these results show that PRMT1, the main cellular methyltransferase in eukaryotic cells [18], recognizes most if not all potential methylation sites of the EWS protein.

In contrast to GST-PRMT1, GST-PRMT3 methylated GST-EWS protein less but more selectively. Mainly the C-terminal RGG box 3 was found to be methylated almost completely whereas the sites in the other two RGG boxes with similar methylation consensus sequences were more or less unmodified, even when the reaction time was prolonged or more activity units were used (data not shown). Thus, the consensus sequences seem not to be sufficient criteria for methylation by GST-PRMT3. Rather structural determinants seem to prevent methylation by GST-PRMT3 in some regions, although they are accessible for GST-PRMT1.

Differential methylation activity of GST-PRMT3 and GST-PRMT1 has already been reported [35]. PRMT3 was shown to methylate undefined substrates distinct from those of PRMT1. Incubation of a hypomethylated yeast rmt1 extract with either PRMT1 or PRMT3 led to the formation of different methylated products [11]. A regulatory effect on the activity of PRMT3 might arise from the zinc-finger domain, which among all PRMT is only found in PRMT3. High concentrations of zinc (500 μM) have been reported to inhibit the methylation of a recombinant arginine-glycine rich protein (GST-GAR) and hypomethylated RAT1 cell extracts by GST-PRMT3. The deletion of the zinc-finger domain in GST-PRMT3 results in complete loss of the methylation activity [36]. In our study, however, addition of neither zinc nor EDTA changed the methylation characteristics of GST-PRMT3 towards the peptide (data not shown). Zinc-finger motifs are not only known as modules that trigger protein/DNA or protein/RNA interactions but they also interact with each other. C_2H_2 type zinc-fingers, the type that is present in PRMT3, are supposed to be capable of interacting with C_2C_2 type zinc-fingers [37]. The zinc-finger domain of PRMT3 might be important to recognize its specific protein or protein-RNA targets for methylation. The EWS protein indeed contains a C_2C_2 type zinc-finger motif and seems to be capable to bind RNA in its RBD. The interaction between the zinc-fingers of PRMT3 and RNA-liganded or RNA-free EWS protein might facilitate the arginine methylation of EWS protein. We cannot confirm an influence of zinc since methylation of GST-EWS protein with GST-PRMT3 in addition of zinc or EDTA was unchanged (data not shown). Recent findings have shown that arginine methylation in vivo is a complex process, which requires not only a methyltransferase and a substrate but also several cofactors. It was postulated that the type II protein arginine N-methyltransferase JBP1 associates with pICln and Sm proteins and symmetrically dimethylates the RG domain of Sm proteins in a 20S methyltransferase complex, the methylosome [38]. The absence of the cofactors in the in vitro reactions could thus be responsible for the reduced methylation of GST-EWS protein by GST-PRMT3.

CARM1/PRMT4, also a type I methyltransferase, was not considered in this study, because the methylated arginine residues of the known substrates like histone H3 [12] or the CREB binding protein [39] are not within an RGG consensus sequence. By methylating the yeast protein Npl3 and GST-GAR, PRMT6 showed a similar substrate specificity like PRMT1 [13]. In our study, however, GST-PRMT6 showed almost no activity towards the reference peptide (1 μM peptide was incubated with 0.1 μM GST-PRMT6) and did not methylate the GST-EWS protein (2 μM GST EWS was incubated with 1 μM GST-PRMT6; data not shown).

In conclusion, we demonstrated that GST-EWS protein is a new efficient substrate for GST-PRMT1 but less for GST-PRMT3 in vitro. MS analysis of the methylation process detected specific methylation sites of PRMT3, which would not be detectable easily in radiolabeling experiments. Most likely also endogenous EWS protein is a strong substrate for PRMT1 but less for PRMT3 in vivo as shown by the pull-down experiments. Up to now, only methylated endogenous EWS was found in cells, be it in the nucleus, the cytoplasm, or in the membrane fraction [20]. Unmethylated sites of the EWS protein seem to be rare suggesting that EWS protein is methylated immediately after translation, but where and by what PRMT cannot be answered yet. It could well be that PRMT3, which is located in the cytosol [11], partly methylates EWS protein first and PRMT1 is completing the process in the nucleus. Our present findings that nuclear EWS strongly interacts with PRMT1 suggest either that PRMT1 shows affinity towards methylated EWS protein or certain sites of nuclear EWS protein might be still unmethylated. This model fits only, if PRMT1 would be localized in the nucleus, as it seems the case for RAT1 cells [11]. However, in HeLa cells, PRMT1 was also shown to be located in both, the nucleus and the cytosol [13,40]. Thus, the place and the sequence of the methylation process of the EWS protein remain speculative at the present state of investigation as it is whether methylation of EWS protein affects its subcellular localization, as reported for proteins like hnRNP [41] or RNA helicase A [42].

Acknowledgements

We are grateful to Dr. Serge Gisler from the Institute of Physiology, University of Zurich, for his kind support in cloning and expression of EWS protein. We would like to thank Doris Grossenbacher and Rouzanna Zakaryan for their support in purification of the recombinant proteins, and Dr. Peter Gehrig (FGCZ) for assistance in the RP-HPLC and MALDI-TOF MS/MS experiments. We also thank Dr. Michael Hottiger and Nazim El-Andaloussi for providing GST-PRMT6, and Dr. Philipp Christen for helpful discussions and critical reading of the manuscript. This work was supported by Swiss National Science Foundation Grant No. 31-66954.01.

Abbreviations: EWS, Ewing Sarcoma protein; AdoMet, S-adenosyl-L-methionine; AdoHcy, S-adenosyl-L-homocysteine; PRMT, protein arginine methyltransferase; GST, glutathione-S-transferase; CARM, coactivator associated arginine N-methyltransferase; JBP, Janus kinase binding protein; MALDI-TOF MS, matrix-assisted laser desorption ionisation time-of-flight

mass spectrometry; r.i., relative intensity; RP-HPLC, reverse phase-high performance liquid chromatography.

References

1. Lee HW, Kim S, Paik WK. S-adenosylmethionine: protein-arginine methyltransferase. Purification and mechanism of the enzyme. Biochemistry 1977;16(1):78-85.
2. Flach J, Bossie M, Vogel J, Corbett A, Jinks T, Willins DA, Silver PA. A yeast RNA-binding protein shuttles between the nucleus and the cytoplasm. Molecular & Cellular Biology 1994;14(12):8399-407.
3. Shen EC, Henry MF, Weiss VH, Valentini SR, Silver PA, Lee MS. Arginine methylation facilitates the nuclear export of hnRNP proteins. Genes & Development 1998;12(5):679-91.
4. Mowen K, David M. Response to matters arising. Cell 2004;119(5):589-90.
5. Meissner T, Krause E, Lodige I, Vinkemeier U. Arginine methylation of STAT1: a reassessment. Cell 2004;119(5):587-9; discussion 589-590.
6. Rajpurohit R, Paik WK, Kim S. Effect of Enzymatic Methylation of Heterogeneous Ribonucleoprotein Particle A1 on Its Nucleic-Acid Binding and Controlled Proteolysis. Biochemical Journal 1994;304:903-909.
7. Klein S, Carroll JA, Chen Y, Henry MF, Henry PA, Ortonowski IE, Pintucci G, Beavis RC, Burgess WH, Rifkin DB. Biochemical analysis of the arginine methylation of high molecular weight fibroblast growth factor-2. Journal of Biological Chemistry 2000;275(5):3150-7.
8. Gary JD, Clarke S. RNA and protein interactions modulated by protein arginine methylation. Progress in Nucleic Acid Research & Molecular Biology 1998;61:65-131.
9. Gary JD, Lin WJ, Yang MC, Herschman HR, Clarke S. The predominant protein-arginine methyltransferase from Saccharomyces cerevisiae. J Biol Chem 1996;271(21):12585-94.
10. Lin WJ, Gary JD, Yang MC, Clarke S, Herschman HR. The mammalian immediate-early TIS21 protein and the leukemia-associated BTG1 protein interact with a protein-arginine N-methyltransferase. J Biol Chem 1996;271(25):15034-44.
11. Tang J, Gary JD, Clarke S, Herschman HR. PRMT 3, a type I protein arginine N-methyltransferase that differs from PRMT1 in its oligomerization, subcellular localization, substrate specificity, and regulation. Journal of Biological Chemistry 1998;273(27):16935-45.
12. Chen D, Ma H, Hong H, Koh SS, Huang SM, Schurter BT, Aswad DW, Stallcup MR. Regulation of transcription by a protein methyltransferase. Science 1999;284(5423):2174-7.
13. Frankel A, Yadav N, Lee J, Branscombe TL, Clarke S, Bedford MT. The novel human protein arginine N-methyltransferase PRMT6 is a nuclear enzyme displaying unique substrate specificity. J Biol Chem 2002;277(5):3537-43.
14. Branscombe TL, Frankel A, Lee JH, Cook JR, Yang Z, Pestka S, Clarke S. PRMT5 (Janus kinase-binding protein 1) catalyzes the formation of symmetric dimethylarginine residues in proteins. Journal of Biological Chemistry 2001;276(35):32971-6.
15. Lee JH, Cook JR, Yang ZH, Mirochnitchenko O, Gunderson S, Felix AM, Herth N, Hoffmann R, Pestka S. PRMT7: A new protein arginine methyltransferase that synthesizes symmetric dimethylarginine. J Biol Chem 2004.
16. Baldwin GS, Carnegie PR. Specific enzymic methylation of an arginine in the experimental allergic encephalomyelitis protein from human myelin. Science 1971;171(971):579-81.
17. Friesen WJ, Paushkin S, Wyce A, Massenet S, Pesiridis GS, Van Duyne G, Rappsilber J, Mann M, Dreyfuss G. The methylosome, a 20S complex containing JBP1 and pICln, produces dimethylarginine-modified Sm proteins. Molecular & Cellular Biology 2001;21(24):8289-300.
18. Tang J, Frankel A, Cook RJ, Kim S, Paik WK, Williams KR, Clarke S, Herschman HR. PRMT1 is the predominant type I protein arginine methyltransferase in mammalian cells. J Biol Chem 2000;275(11):7723-30.
19. Bachand F, Silver PA. PRMT3 is a ribosomal protein methyltransferase that affects the cellular levels of ribosomal subunits. Embo J 2004;23(13):2641-50.
20. Belyanskaya LL, Gehrig PM, Gehring H. Exposure on cell surface and extensive arginine methylation of ewing sarcoma (EWS) protein. Journal of Biological Chemistry 2001;276(22):18681-18687.
21. Arvand A, Denny CT. Biology of EWS/ETS fusions in Ewing's family tumors. Oncogene 2001;20(40):5747-5754.

22. Plougastel B, Zucman J, Peter M, Thomas G, Delattre O. Genomic structure of the EWS gene and its relationship to EWSR1, a site of tumor-associated chromosome translocation. Genomics 1993;18(3):609-15.
23. Delattre O, Zucman J, Plougastel B, Desmaze C, Melot T, Peter M, Kovar H, Joubert I, Dejong P, Rouleau G and others. Gene Fusion with an Ets DNA-Binding Domain Caused by Chromosome-Translocation in Human Tumors. Nature 1992;359(6391):162-165.
24. Rossow KL, Janknecht R. The Ewing's sarcoma gene product functions as a transcriptional activator. Cancer Research 2001;61(6):2690-2695.
25. Li KKC, Lee KAW. Transcriptional activation by the Ewing's sarcoma (EWS) oncogene can be cis-repressed by the EWS RNA-binding domain. Journal of Biological Chemistry 2000;275(30):23053-23058.
26. Kzhyshkowska J, Schutt H, Liss M, Kremmer E, Staubert R, Wolf H, Dobner T. Heterogeneous nuclear ribonucleoprotein E1B-AP5 is methylated in its Arg-Gly-Gly (RGG) box and interacts with human arginine methyltransferase HRMT1L1. Biochemical Journal 2001;358:305-314.
27. Weighardt F, Biamonti G, Riva S. Nucleocytoplasmic Distribution of Human Hnrnp Proteins - a Search for the Targeting Domains in Hnrnp A1. Journal of Cell Science 1995;108:545-555.
28. Zamore PD, Patton JG, Green MR. Cloning and Domain-Structure of the Mammalian Splicing Factor U2af. Nature 1992;355(6361):609-614.
29. Krainer AR, Mayeda A, Kozak D, Binns G. Functional Expression of Cloned Human Splicing Factor Sf2 - Homology to Rna-Binding Proteins, U1-70k, and Drosophila Splicing Regulators. Cell 1991;66(2):383-394.
30. Belyanskaya LL, Delattre O, Gehring H. Expression and subcellular localization of Ewing sarcoma (EWS) protein is affected by the methylation process. Exp Cell Res 2003;288(2):374-81.
31. Zhen Y, Xu N, Richardson B, Becklin R, Savage JR, Blake K, Peltier JM. Development of an LC-MALDI method for the analysis of protein complexes. J Am Soc Mass Spectrom 2004;15(6):803-22.
32. Gehrig PM, Hunziker PE, Zahariev S, Pongor S. Fragmentation pathways of N(G)-methylated and unmodified arginine residues in peptides studied by ESI-MS/MS and MALDI-MS. J Am Soc Mass Spectrom 2004;15(2):142-9.
33. Rappsilber J, Friesen WJ, Paushkin S, Dreyfuss G, Mann M. Detection of arginine dimethylated peptides by parallel precursor ion scanning mass spectrometry in positive ion mode. Anal Chem 2003;75(13):3107-14.
34. Merrill BM, Lopresti MB, Stone KL, Williams KR. Amino acid sequence of UP1, an hnRNP-derived single-stranded nucleic acid binding protein from calf thymus. Int J Pept Protein Res 1987;29(1):21-39.
35. Liu Q, Dreyfuss G. In vivo and in vitro arginine methylation of RNA-binding proteins. Molecular & Cellular Biology 1995;15(5):2800-8.
36. Frankel A, Clarke S. PRMT3 is a distinct member of the protein arginine N-methyltransferase family. Conferral of substrate specificity by a zinc-finger domain. Journal of Biological Chemistry 2000;275(42):32974-82.
37. Mackay JP, Crossley M. Zinc fingers are sticking together. Trends in Biochemical Sciences 1998;23(1):1-4.
38. Brahms H, Meheus L, de Brabandere V, Fischer U, Luhrmann R. Symmetrical dimethylation of arginine residues in spliceosomal Sm protein B/B' and the Sm-like protein LSm4, and their interaction with the SMN protein. Rna-A Publication of the Rna Society 2001;7(11):1531-42.
39. Chevillard-Briet M, Trouche D, Vandel L. Control of CBP co-activating activity by arginine methylation. EMBO Journal 2002;21(20):5457-66.
40. Cote J, Boisvert FM, Boulanger MC, Bedford MT, Richard S. Sam68 RNA binding protein is an in vivo substrate for protein arginine N-methyltransferase 1. Mol Biol Cell 2003;14(1):274-87.
41. Xu C, Henry MF. Nuclear Export of hnRNP Hrp1p and Nuclear Export of hnRNP Npl3p Are Linked and Influenced by the Methylation State of Npl3p. Mol Cell Biol 2004;24(24):10742-56.
42. Smith WA, Schurter BT, Wong-Staal F, David M. Arginine methylation of RNA helicase a determines its subcellular localization. J Biol Chem 2004;279(22):22795-8.

4. Identification of proteins interacting with Protein Arginine Methyl-transferase 8: The Ewing Sarcoma (EWS) Protein binds independent of its methylation state.

Steffen Pahlich, Rouzanna P Zakaryan, and Heinz Gehring

Submitted for publication

Summary

Protein arginine methylation is a eukaryotic post-translational modification that plays a role in transcription, mRNA splicing and transport, in protein-protein interaction, and cell signaling. The type I protein arginine methyltransferase (PRMT) 8 is the only member of the human PRMT family that is localized at the cell membrane and its endogenous substrates have remained unknown as yet.

In this study, we identified more than 20 PRMT8-binding partners in pull-down experiments using recombinant PRMT8 as bait followed by mass spectrometric identification of the bound proteins. Among the extracted proteins were several heterogeneous nuclear ribonucleoproteins (hnRNP), RNA-helicases (DEAD box proteins), the TET-family proteins TLS, EWS, and $TAF_{II}68$, and caprin, which all contain RGG motifs, the preferred methylation sequence. They therefore are potential substrates of PRMT8. Further identified proteins were actin, tubulin, and heat shock proteins.

We characterized the interaction between PRMT8 and the EWS protein in more detail. Although binding of endogenous and recombinant EWS protein to PRMT8 as well as co-localization in HEK cells was observed, in vitro methylation assays revealed a rather poor methyltransferase activity of PRMT8 towards the EWS protein and a synthetic RGG-rich reference peptide in comparison to PRMT1. The main interaction site of the EWS protein with PRMT8 was determined to be the C-terminal RGG box 3. Remarkably, the complete methylation of the EWS protein did not abrogate the binding to PRMT8, pointing to a further adapter role of PRMT8 at the cell membrane.

Introduction

Protein arginine methylation is a eukaryotic post-translational modification catalyzed by protein arginine methyltransferases (PRMTs) (1). Currently, nine different human PRMTs are known, which all share a highly conserved S-adenosyl-L-methionine (AdoMet) binding site and a less conserved substrate binding site. They are classified in type I or type II enzymes, that form N^G-monomethylarginine (MMA) and asymmetric ω-N^G,N^G-dimethylarginine (ADMA), or MMA and symmetric ω-N^G,$N^{G`}$-dimethylarginine (SDMA), respectively. The PRMTs differ in their subcellular localizations, N- and C-terminal extensions, and their substrate specificity. Typical substrates of PRMTs are nuclear DNA- or RNA-binding proteins like heterogenous nuclear ribonucleoproteins (hnRNPs), spliceosomal proteins, and histones. Arginine methylation plays a role in processes like transcriptional control, mRNA splicing and -transport, protein-nucleic acids and protein-protein interactions, as well as cell signaling (for reviews, see (2,3)). However, the list of identified methylated proteins is still growing and for many of them, the role of their methylation is largely unknown.

Human PRMT8 is the type I methyltransferase with the highest sequence similarity to PRMT1, the dominant methyltransferase in human cells. The enzyme seems to be expressed only in brain tissue and was shown, to associate with the cell membrane upon myristoylation (4). The type I methyltransferase activity of recombinant GST-PRMT8 was demonstrated towards histone H4 and two recombinant RGG-rich proteins (GST-GAR, GST-Npl1), but proved much lower than that of GST-PRMT1 (4). The endogenous substrate proteins of PRMT8 as well as its role at the cell membrane are still unknown.

The human Ewing sarcoma (EWS) protein is highly methylated (5) and belongs to the TET-family (TLS/FUS, EWS, $TAF_{II}68$) of RNA-binding proteins. It acts as a transcriptional coactivator (6,7) and might play a role in splicing or mRNA transport (8). EWS protein-deficient mice show a high percentage (~90%) of post-natal mortality and the few surviving animals fail to develop intact B-lymphocytes and have defects in chromosome pairing during meiosis (9). The EWS protein is localized mainly in the nucleus, but minor fractions were also detected in the cytosol and associated with the plasma membrane (5,10). Chromosomal translocations in the family of Ewing tumors lead to the expression of chimeric fusion proteins between the N-terminal transcriptional activation domain (EAD) of the EWS protein and DNA-binding domains of different ETS (erythroblastosis virus-transforming sequence) transcription factors (11). In the wild-type protein, such transcriptional activity of the

oncogenic EAD is repressed by the C-terminal RNA-binding domain (RBD) (12). The RBD consists of an RNA-recognition motif (RRM), a putative C_2C_2 zinc finger, and three arginine-glycine rich regions (RGG boxes 1-3). All 30 arginines within these RGG boxes were found to be asymmetrically dimethylated in vitro and in vivo (5,13).

In the present study we screened for binding partners and potential substrates of PRMT8 to investigate the role of this membrane-bound methyltransferase. With GST pull-down experiments followed by mass spectrometric identification of the bound proteins we could show that several nuclear proteins like hnRNPs, the TET family proteins TLS/FUS, EWS, and $TAF_{II}68$, the RNA helicases DDX1 and DDX3X, but also structural proteins like actin and tubulin bind to PRMT8. We further characterized the interaction between PRMT8 and the EWS protein, which is also associated with the cell membrane, and it turned out, that the methyltransferase activity of PRMT8 towards the EWS protein is relatively poor. The interaction, however, was maintained independent of the methylation of the EWS protein, although its RGG box 3 was determined to be the main interaction site with PRMT8, indicating an additional cellular role of PRMT8 than just to act as methyltransferase.

Materials and Methods

Materials – The construct pGEX(SN)-PRMT1 was a generous gift from Dr. Harvey R. Herschman (UCLA, Molecular Biology Institute, Los Angeles). GST-PRMT8 (pGEX-6P-1-vector) and GFP-PRMT8 (pEGFP-N1-vector) were kindly provided by Dr. Mark T. Bedford (University of Texas M. D. Anderson Cancer Center, Smithville, Texas). The cDNA of EWS was kindly provided by Dr. Olivier Delattre (Institut Curie, Pathologie Moléculaire des Cancers, Paris Cedex).

Expression and purification of GST-fusion proteins – GST-PRMT1, GST-PRMT8, and GST-EWS were expressed in E.coli XL1-Blue competent cells and purified as described (4,13). Each purified protein showed a single major band after SDS-PAGE and Coomassie blue staining.

Expression of His-GFP-EWS-myc-His fusion proteins - *EWS* cDNA was amplified by PCR using XhoI-NotI-forward and EcoRI-reverse primers 5'-TCTCTCGAGCGGCCGCGCCACCATGGCGTCCACGGATTACAGTACC-3' and 5'-CCGAATTCGTAGGGCCGATCTCTGCGCTCCTG-3', respectively. The PCR product was treated with restriction enzymes XhoI and EcoRI and subcloned into pcDNA3.1(-)B/myc-His

(Invitrogen) to generate in-frame fusion. To construct the pEGFP-N2-EWS-myc-His vector, the EWS-myc-His sequence was amplified by PCR using the above-mentioned XhoI-NoI-forward primer and SacII-His-reverse primer- 5'-TCCCCGCGGGGATGATGATGATGATGATGGTCGACG-3', and pcDNA3.1(-)B-EWS-myc-His expression vector as a template. The resulting fragment was digested with XhoI and SacII restriction enzymes, and incorporated into the pEGFP-N2 plasmid (Clontech). The His-GFP-EWS-myc-His vector was constructed by amplification of the His-GFP fragment from pEGFP-N2-EWS-myc-His using the primers NheI-His-forward (5'-TCTGCTAGCGCCACCATGGCCGTCGACCATCATCATCATCATCAT-3', and NotI-GFP-reverse (5'-TGCGTCGCGGCCGCTCTTGTACAGCTCGTCCATGCCGAG-3') and cloning of the PCR product into the pcDNA3.1(-)B-EWS-myc-His plasmid. The truncated RBD-YFP-constructs (Fig. 4.5 E) were produced as described (14).

Antibodies – Anti-EWS antibody (against the EWS protein N-terminus, used in 1:5000 dilution) were kindly provided by Dr. O. Delattre (Institut Curie, Pathologie Moléculaire des Cancers, Paris Cedex)). Anti-GFP antibody (N-terminal) from Sigma, which recognizes also YFP, was used in a 1:300 dilution, and the anti-His antibody from Sigma (1:2000 dilution) was a generous gift of Dr. A. Plückthun (Department of Biochemistry, University of Zurich).

Cell culture, preparation of total cell lysates – Jurkat cells were cultivated in RPMI medium at 37°C. To obtain hypomethylated proteins, adenosine-2`,3`-dialdehyde (AdOx, from Sigma) was added in a final concentration of 20 μM to the medium and the cells were further incubated for 24 h at 37°C. Cells were harvested, washed twice in ice cold PBS buffer and lysed for 30 min on ice in lysis buffer A (50 mM Tris, pH 8.0; 200 mM NaCl; 1 mM PMSF; 1 mM DTT; 1% Triton-X 100; protease inhibitor mix (Roche)). The lysate was centrifuged for 30 min at 13000 rpm and the supernatant was designated as total Jurkat lysate. Human embryonic kidney (HEK) 293 (T) cells were grown in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% fetal calf serum (FCS) and 1% (w/v) of penicillin and streptomycin in a humidified 10% CO_2 atmosphere at 37ºC. To obtain hypomethylated HEK cells, AdOx was added in a final concentration of 20 μM to the medium and the cells were further incubated for 24 h. Cells were harvested, washed twice in ice cold PBS and lysed in lysis buffer B (50 mM Tris, pH 8.0; 200 mM NaCl; 10 mM β–mercaptoethanol; 10% glycerol; 1% Triton-X 100; EDTA-free protease inhibitor mix (Roche)) for 30 min on ice. After centrifugation for 30 min at 13000 rpm, the obtained supernatant was designated as total HEK lysate.

Transfection, expression, and purification of EWS constructs in HEK cells – HEK cells were grown to ~50% confluency for transient transfection through the calcium phosphate precipitation method. Expression of the transfected proteins was checked after 20 h by fluorescence microscopy. For purification of the EWS constructs, the transfected cells were lysed as described above, 10 µl His-SELECT Nickel affinity gel (50% slurry, Sigma) was added to 500 µl of the total HEK lysate and incubated for two h at room temperature (RT) under constant shaking. The affinity gel was washed twice with lysis buffer B containing 10 mM imidazole and the bound proteins were eluted specifically using 50 µl of lysis buffer B and 250 mM imidazole for 30 min at RT under constant shaking.

Immunocytochemistry **-** For protein visualization analysis, HEK cells were grown on glass coverslips for 24 h to 40% confluency. For the co-localization assay of the EWS protein and PRMT8, the cells were transiently transfected with plasmid expressing PRMT8-GFP (4) through calcium phosphate precipitation method. The cells were washed 24 h after transfection with PBS, fixed and permeabilized with 4% paraformaldehyde (PFA) and 1% Triton X-100 for 10 min followed by washing and addition of blocking solution (10% FCS and 0.1% glycine in PBS) for 1 h. The cells were incubated first with rabbit anti-C-terminal EWS antibody (1:100 dilution, kindly provided by Dr. O. Delattre (Institut Curie, Pathologie Moléculaire des Cancers, Paris Cedex)) for 1 h at RT, followed by washing and incubation with indocarbocyanine (Cy3)-coupled goat anti-rabbit secondary antibodies (1:500 dilution, kindly provided by Dr. P. Sonderegger, Department of Biochemistry, University of Zurich, Switzerland) for 1 h at RT. The cells were washed, stained with 4´,6-diamidino-2-phenylindol (DAPI,Roche), mounted using Vectashield medium (Vector) onto the glass slides, and analyzed by confocal microscopy as described (14) using excitation-emission wavelengths of 405-470 nm, 488-509 nm and 550-570 nm for DAPI, GFP and Cy3, respectively.

Methylation of the cell extract – Hypomethylated total HEK cell lysate (70 µg of total protein) was incubated with GST-PRMT1 (2 µg) or GST-PRMT8 (2 µg) in the presence of 0.4 µCi *S*-adenosyl-L-[*methyl*-^{3}H]methionine (^{3}H-AdoMet, 60 Ci/mmol from a 1 mCi/ml stock solution) in PBS for 3 h at 30 °C. The reaction was stopped by the addition of two fold Laemmli sample loading buffer and heating at 95°C for five min. The samples were then subjected to SDS-PAGE (10% acrylamide gels), the gel was incubated in NAMP 100 enhancer (Amersham) for 2 h in the dark, the gel was dried and methylated proteins were visualized by fluorography.

GST pull-down – GST or GST-PRMT8 (10 μg) was bound to glutathione-coupled agarose (Sigma) for 2 h at RT and the agarose was washed afterwards twice with 100 mM Na-Borate and twice with 50 mM Tris, pH 8.0. Either total cell lysate (1 mg of total protein) or purified EWS constructs (1-2 μg) was added in the presence of binding buffer (50 mM Tris, pH 8.0; 100 mM NaCl; 1 mM PMSF; 1 mM DTT; 0.1% Triton X-100; protease inhibitor mix (Roche)), and the sample was incubated for one h at RT and one h at 4 °C under constant shaking. The agarose was washed three times with binding buffer and bound proteins were specifically eluted with 20 mM glutathione / 50 mM Tris, pH 8.0. The eluted proteins were separated by SDS-PAGE (10% acrylamide) and analyzed by western blotting or in-gel digestion followed by mass spectrometry.

Peptide methylation by GST-PRMT8 – Peptide methylation was performed as described (13). Briefly, 2 μM of a synthetic peptide (GRGGFGGRGGFRGGRGG-NH_2, from ANAWA Biomedical Services & Products) was incubated with 1 μg GST-PRMT8 and 500 μM *S*-adenosyl-L-methionine (AdoMet) in 50 mM Tris, pH 8.0 for various time periods at 30°C. The methylation reaction was stopped by addition of an equal volume of 50% (v/v) acetonitrile (ACN) / 0.1% (v/v) trifluoroacetic acid (TFA). Chymotryptic digestion of the peptide was performed in solution with a final concentration of 1 μg chymotrypsin/ml for 15 min at 30°C. The determination of the kinetic parameters was performed as described (13).

Methylation of recombinant GST-EWS protein – GST-EWS was incubated with GST-PRMT1 (2 μg) or GST-PRMT8 (2 μg) in the presence of 0.4 μCi ^{3}H-AdoMet (60 Ci/mmol from a 1 mCi/ml stock solution) and 500 μM AdoMet in PBS for 3h at 30 °C. The reaction was terminated by the addition of 10 μl of two fold Laemmli sample loading buffer to 10 μl of the reaction mixture and by heating for 5 min at 95°C. The samples were then subjected to SDS-PAGE (10% acrylamide gels), and incubated radioactivity was detected as described above. For mass spectrometry analysis, the methylation of GST-EWS was performed with unlabeled AdoMet only.

Mass spectrometry (MS) – MS analyses were performed at the Functional Genomics Center Zürich (FGCZ) and at the Protein Analysis Unit of the University of Zürich.

To identify the PRMT8-binding proteins from the GST pull-down, the whole gel lane was cut into slices and each slice was in-gel digested with trypsin. The tryptic peptides were pooled, desalted using C_{18}-ZipTips (Millipore, Bedford, MA), and analyzed by nano-LC-MSMS using an LTQ-FT (Thermo Electron, Bremen, Germany). Peptides were separated on

a nano-HPLC (Eksigent Technologies, Dublin CA) online prior to MS analysis on a home made C_{18} reversed phase column (Magic 3 µm, 100 Å C18 AQ, Michrom, Auburn, CA) using an acetonitrile/water system at a flow rate of 200 nl/min. Tandem mass spectra were acquired in a data-dependent manner. Typically, 5 MS/MS were performed after each high accuracy spectral acquisition range survey. A customized database (38246 entries) comprising the human portion (taxonomy ID: 9606) of the UniProt database (http://www.ebi.ac.uk) was interrogated using the Mascot search algorithm (15) on an in-house server (FGCZ). For the search, three missed trypsin cleavages were allowed. The precursor ion tolerance and fragment ion tolerance were set to 2.5 ppm and 0.6 Da, respectively, and oxidized methionine was set as variable modification. Only proteins, which were identified in at least two independent pull-down experiments were counted as potential interaction partner of PRMT8. The MSMS spectra of proteins, whose identification was based only on one or two assigned peptides, were checked manually for a correct peak assignment to exclude false positives.

The methylation time course of the peptide was analyzed by MALDI-TOF-MS on a Bruker Biflex instrument as described (13). The MSMS analysis of the digested partially methylated peptide was performed on a Bruker-Daltonics Ultraflex TOF/TOF II MALDI mass spectrometer operated in the positive ion reflector mode with 150 ns as delayed extraction time. The nitrogen laser (337 nm) was set to a repetition rate of 50 Hz and the ion acceleration voltage was 25 kV. The spectra were calibrated externally.

The in vitro methylated EWS protein was in-gel digested with trypsin or chymotrypsin, the peptides were desalted using C_{18}-ZipTips (Millipore, Bedford, MA) and analyzed by LC-MALDI-TOF/TOF-MS as described (13).

Results

GST pull-down with PRMT8

To identify endogenous interaction partners and potential substrate proteins of PRMT8, GST pull-down experiments followed by mass spectrometric analysis of the bound proteins were carried out. In the cell, methyl-accepting proteins, the potential substrates of PRMT8, are usually already methylated (16) and might not bind anymore to the PRMTs. Unmethylated substrate proteins (hypomethylated cell lysate) were obtained by treating Jurkat cells with the methyltransferase inhibitor AdOx (adenosine-2´,3´-dialdehyde)(16). Hypomethylated total

lysate was added to recombinant GST-PRMT8 bound to glutathione-coupled agarose. After extensive washing, GST-PRMT8 and its bound interaction partners were specifically released with glutathione from the agarose and the extracted proteins were resolved by SDS-PAGE (Fig. 4.1). In addition to the dominant GST-PRMT8 band at 66 kDa and the GST band at 30 kDa (due to partial degradation of GST-PRMT8), several co-purifying proteins were extracted; no dominant protein bands except GST are present if the bait was recombinant GST. For identification of the bound proteins, the whole gel lane was cut into slices, each slice was in-gel digested with trypsin, the extracted peptides were pooled and analyzed by LC-MSMS as described in experimental procedures. Only proteins, which were identified in at least two independent pull-down experiments and were not present in any GST-control experiment were counted as potential interaction partners of PRMT8.

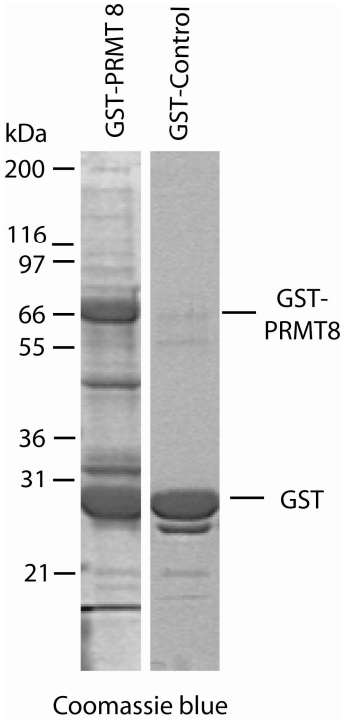


Figure 4.1: GST pull-down with PRMT8 as bait. Hypomethylated total cell lysate from Jurkat cells was used for a pull-down experiment with either recombinant GST-PRMT8 or recombinant GST as bait. Extracted proteins were resolved by SDS-PAGE and visualized by Coomassie blue staining of the gel. Molecular mass markers in kDa are indicated.

More than 20 different proteins were identified by mass spectrometry (Table 1). Because PRMT1 and PRMT8 are known to form homo- and heterodimers (4), the identification of endogenous PRMT1, pulled down by PRMT8, was expected. Among the additional identified interaction partners are the three members of the TET-family of RNA-binding proteins (TLS/FUS, EWS, and $TAF_{II}68$), the heterogeneous nuclear ribonucleoproteins (hnRNPs) K, U, H3, and A3, the RNA helicases DDX1 and DDX3X, and caprin, which all have RGG motifs within their amino acid sequence and thus might be potential substrate proteins of PRMT8. But also structural proteins like actin and tubulin, as well as heat shock proteins were identified.

Table 1: Identified proteins of the GST-PRMT8 pull-down. Only proteins identified in at least two independent experiments are listed with their corresponding mascot scores. The molecular mass (kDa) as well as the presence of arginine-glycine rich regions (RGG) within the protein sequece is indicated.

(primary accession number) protein name	kDa	Mascot score exp. 1	Mascot score exp. 2	RGG
(Q99873) Protein arginine N-methyltransferase 1	41.5	521	378	
(P35637) RNA-binding protein TLS/FUS	53.4	586	246	yes
(Q01844) RNA-binding protein EWS	68.5	498	118	yes
(Q92804) RNA-binding protein TAF(II)68	61.8	882	217	yes
(P51991) Heterogeneous nuclear ribonucleoprotein A3	39.6	390	24	yes
(P31942) Heterogeneous nuclear ribonucleoprotein H3	36.9	498	307	yes
(P61978) Heterogeneous nuclear ribonucleoprotein K	50.9	565	1023	yes
(Q9BQ09) Heterogeneous nuclear ribonucleoprotein U	90.5	783	200	yes
(Q92499) ATP-dependent RNA helicase DDX1 (DEAD box protein 1)	82.4	463	140	
(O00571) ATP-dependent RNA helicase DDX3X (DEAD box protein 3)	73.2	763	38	yes
(Q9BV09) Caprin (Cytoplasmic activation/proliferation-associated protein 1)	116	423	59	yes
(P60709) Actin, cytoplasmic 1 (Beta-actin)	41.7	782	475	
(Q9NY65) Tubulin alpha-8 chain (Alpha-tubulin 8)	50.1	110	352	
(P68363) Tubulin alpha-ubiquitous chain (Tubulin K-alpha-1)	50.2	341	168	
(P07437) Tubulin beta-2 chain	49.6	475	18	
(P38646) Stress-70 protein, mitochondrial precursor	73.7	1607	103	
(P08107) Heat shock 70 kDa protein 1	70	425	165	
(Q9HAV7) GrpE protein homolog 1, mitochondrial precursor	24.3	264	38	
(P11586) C-1-tetrahydrofolate synthase, cytoplasmic	101.6	519	151	
(P68104) Elongation factor 1-alpha 1	50.1	322	286	
(Q8WXH0) Nesprin-2 (Nuclear envelope spectrin repeat protein 2)	796.4	137	29	
(P31327) Carbamoyl-phosphate synthase, mitochondrial precursor (EC 6.3.4.16)	164.9	63	58	

In vitro methylation of the EWS protein by PRMT8

The EWS protein is known to be methylated within its RGG-boxes (13) and beside its nuclear localization, the EWS protein was found to be associated to the cell membrane (5). Therefore, we investigated whether PRMT8 is able to methylate the EWS protein in vitro. Recombinant GST-EWS protein was incubated in the presence of [^{3}H]-AdoMet (*S*-adenosyl-L-[*methyl*-^{3}H]methionine) with either GST-PRMT8 or GST-PRMT1 for 3 h, and the incorporation of radiolabeled methyl groups was detected by autoradiography (Fig. 4.2 A). A clear band of methylated EWS protein could be seen in the presence of PRMT1, however only a faint signal was detected with PRMT8 indicating a very weak methyltransferase activity towards the EWS protein. To verify these results and to identify possible methylation sites, the EWS protein was methylated by PRMT8 with unlabeled AdoMet and after separation of the proteins by SDS-PAGE the EWS protein was in-gel digested either with

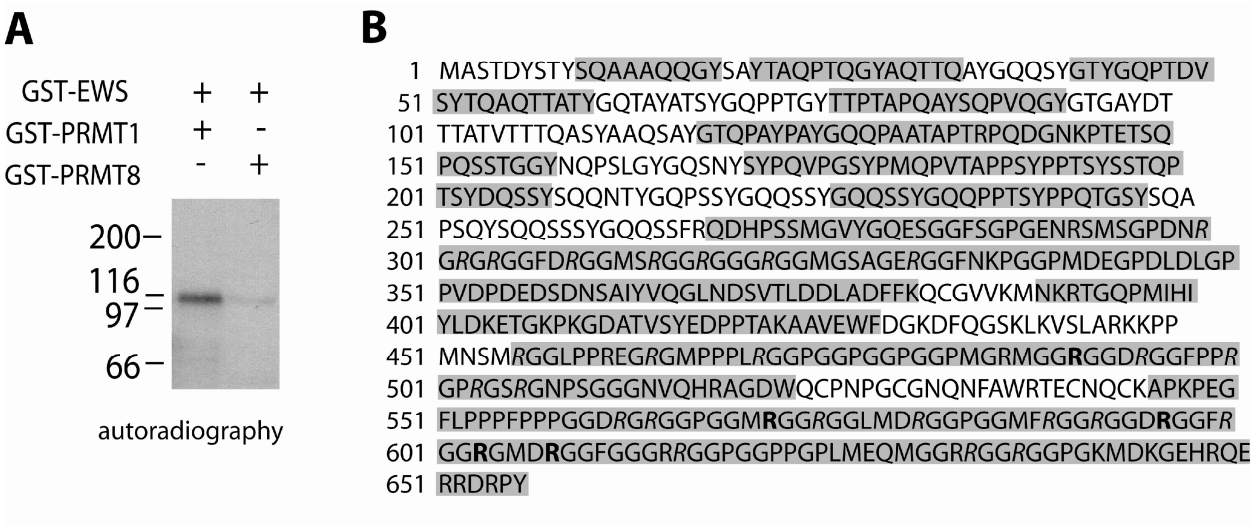


Figure 4.2: In-vitro methylation of the EWS protein by PRMT8. **(A)** GST-EWS was incubated either with GST-PRMT1 or GST-PRMT8 in the presence of ^{3}H-AdoMet for 3h at 30°C, the proteins were resolved by SDS-PAGE and radioactivity incorporated in the recombinant EWS-protein (~110 kDa) was visualized by autoradiography after five days of exposure at -80°C. **(B)** Sequence coverage of the mass spectrometric analysis of the EWS protein, methylated by GST-PRMT8. The EWS protein band was excised from the gel, in-gel digested with either trypsin or chymotrypsin, and the resulting peptides were analyzed by MALDI-TOFTOF-MS. Both, the tryptic and chymotryptic peptides cover a large part of the EWS protein amino acid sequence (shaded), including all 30 potential methylation sites (3,5). The five arginines, which were found to be asymmetrically dimethylated, are shown in bold (**R**), the 25 unmethylated potential methylation sites are in italic (*R*).

trypsin or chymotrypsin and the resulting peptides were analyzed by MALDI-TOFTOF-MS. All 30 potential methylation sites were covered by the peptides, but only five arginines (R490, R572, R596, R603, and R607) were found to be dimethylated (Table 2; Fig. 4.2 B). The appearance of the fragment ion at m/z 46 (dimethylamin) in their MSMS spectra confirmed the asymmetric dimethylation of these arginines (data not shown) (17,18). Three of them (R572, R596, R607) were also present in monomethylated or unmethylated form. All other potential methylation sites were found to be unmodified (Table 2).

Methyltransferase activity of PRMT8 towards a peptide

PRMT8 recognized only five arginine residues of the EWS protein as substrate even though all 30 arginines, that are known to be methylated within the EWS protein (5,13) are situated in highly similar RG(G)-motifs. To test the methylation characteristics and activity of GST-PRMT8 towards RGG motifs in vitro, we used a synthetic peptide containing four arginine methylation sites (GRGGFGGRGGFRGGRGG-NH_2) as substrate and analyzed the process of methylation as a function of time by mass spectrometry (Fig 4.3 A). This method has been proven to be very useful allowing even a quantitative analysis of peptide methylation (13).

Table 2: Identified EWS peptides, which cover all 30 potential methylation sites of the EWS protein. Monomethylated arginines are shown in bold (**R**), asymmetric dimethylated arginines are bold and underlined (**R**). Five arginines were found to be dimethylated, three of them also mono- or unmethylated, whereas 25 arginines were found to be unmodified.

Calc. Mass	Obsrv. Mass	± Da	Start Seq.	End Seq.	Sequence	Ion Score	Modification	Ion at m/z 46
tryptic peptides								
915.46	915.45	-0.01	566	575	GGPGGM**R**GGR	22	1 Methyl (R)	
929.47	929.48	0.01	566	575	GGPGGM**R**GGR	21	1 di-Methylation (R)	yes
778.37	778.36	0.00	582	589	GGPGGMFR	41		
835.42	835.41	0.00	593	600	GGD**R**GGFR	12	1 Methyl (R)	
849.43	849.43	0.00	593	600	GGD**R**GGFR	18	1 di-Methylation (R)	yes
1080.50	1080.52	0.02	604	614	GMD**R**GGFGGGR	24	1 Methyl (R)	
1392.69	1392.69	-0.01	601	614	GG**R**GMD**R**GGFGGGR	6	2 di-Methylation (R)	yes
1750.85	1750.86	0.02	615	632	RGGPGGPPGPLMEQMGGR	64		
chymotryptic peptides								
1773.80	1773.83	0.02	286	302	SGPGENRSMSGPDNRGR	53		
2248.04	2248.06	0.02	286	307	SGPGENRSMSGPDNRGRGRGGF	35		
796.34	796.34	0.00	305	312	GGFDRGGM	43		
1039.47	1039.47	0.00	305	314	GGFDRGGMSR	20		
1679.78	1679.78	0.01	315	333	GGRGGGRGGMGSAGERGGF	77		
1686.92	1686.95	0.02	455	470	RGGLPPREGRGMPPPL	43		
2087.01	2087.07	0.06	465	487	GMPPPLRGGPGGPGGPGGPMGRM	60		
1338.60	1338.61	0.01	472	487	GGPGGPGGPGGPMGRM	83		
1133.53	1133.56	0.03	481	492	GGPMGRMGG**R**GG	14	di-Methylation(R),Ox.(M)	yes
1595.83	1595.85	0.03	488	503	GGRGGDRGGFPPRGPR	29		
1625.84	1625.86	0.03	491	506	GGDRGGFPPRGPRGSR	20		
1633.79	1633.76	-0.03	501	517	GPRGSRGNPSGGGNVQH	76		
2186.15	2186.21	0.06	545	565	APKPEGFLPPPFPPPGGDRGR	69		
1915.96	1916.00	0.05	552	571	LPPPFPPPGGDRGRGGPGGM	48		
1251.56	1251.58	0.02	576	588	GGLMDRGGPGGMF	58		
1091.55	1091.56	0.02	589	599	RGGRGGDRGGF	46		
1105.56	1105.56	-0.01	589	599	RGGRGGD**R**GGF	17	Methyl (R)	
1119.58	1119.56	-0.02	589	599	RGGRGGD**R**GGF	31	di-Methylation (R)	yes
1165.56	1165.56	0.00	600	610	RGGRGMDRGGF	41		
1807.87	1807.88	0.01	611	629	GGGRRGGPGGPPGPLMEQM	117		
1750.85	1750.89	0.04	615	632	RGGPGGPPGPLMEQMGGR	73		
1538.76	1538.74	-0.02	633	647	RGGRGGPGKMDKGEH	25		

Methylation of the peptide by GST-PRMT8 proceeded very slowly. After one h of incubation, the majority of the peptide was still unmethylated. After seven h, the main peak was the four times methylated peptide (+56 Da) and it was still the dominant signal even after 22 h. Minor fractions of the peptide are also five (+70 Da) and six times (+84 Da) methylated, but a complete methylation of the peptide was not achieved. Neither longer incubation times nor the addition of fresh PRMT8 completed the methylation of the peptide (data not shown). The initial rates at different peptide concentrations indicated a similar K_m of PRMT8 and PRMT1

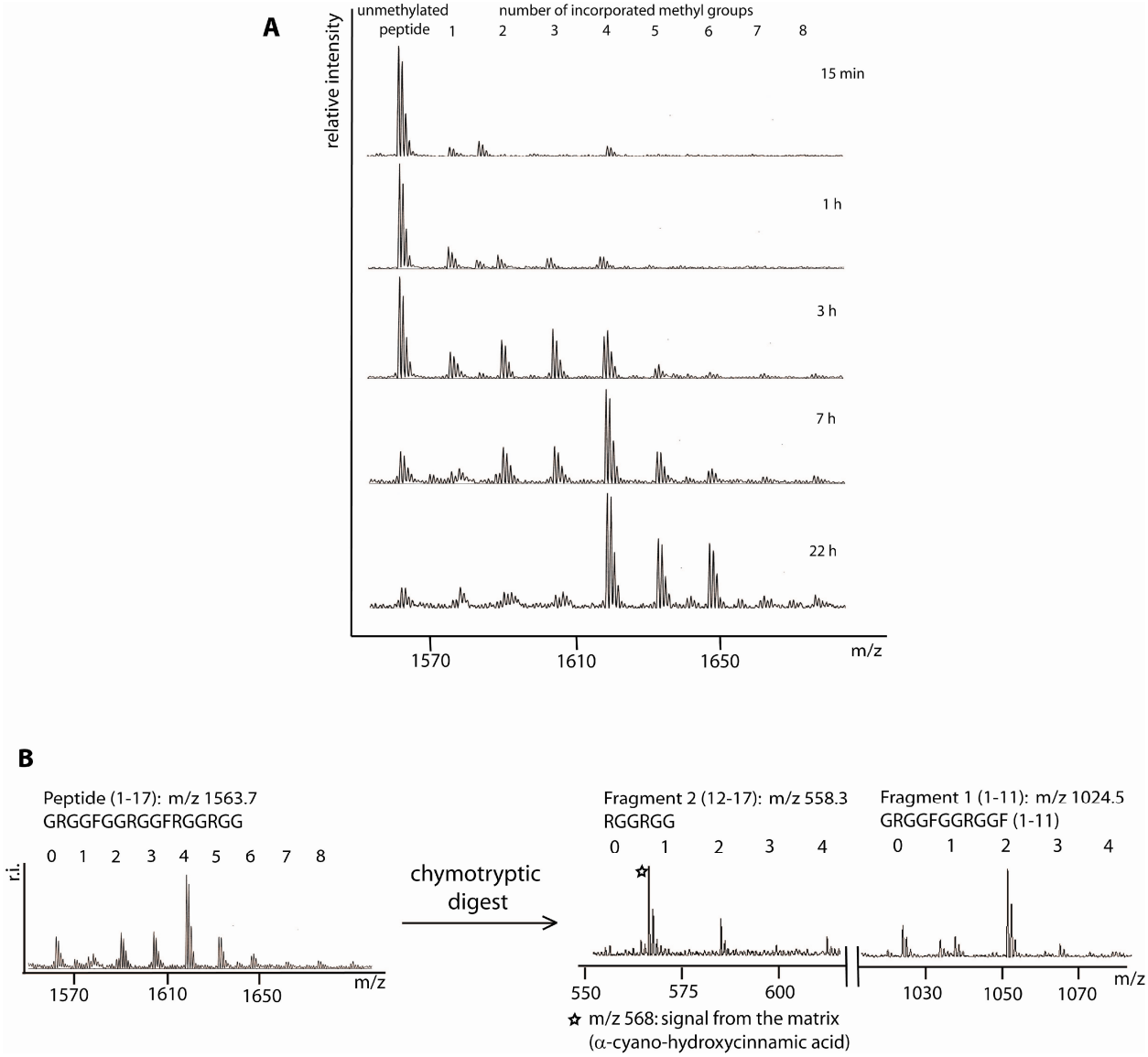


Figure 4.3: Methylation of a synthetic peptide by PRMT8. **(A)** A synthetic peptide (m/z 1563.6, GRGGFGGRGGFRGGRGG-NH_2) with four arginine methylation sites was incubated with PRMT8 and AdoMet at 30°C and analyzed at the indicated time points by MALDI-TOF-MS. Each incorporated methyl group leads to a mass increase of 14 Da. The numbers of incorporated methyl groups are indicated above the peaks. **(B)** Methylation of the peptide by GST-PRMT8 was stopped after seven h, the partly methylated peptide was digested with chymotrypsin, and the methylation state of the resulting fragments (1-11; 12-17) was analyzed by MALDI-TOF-MS. Again, numbers of incorporated methyl groups are indicated. The intensive peak at m/z 568 labeled with an asterisk is a signal arising from the matrix (α-CCA).

(1.3 μM and 0.8 μM, respectively) towards this peptide but the catalytic efficiency (k_{cat}/K_m) was only 1/30 of that of PRMT1 (supplemental Fig. 4.1) (13).

To characterize the methylation pattern, the partly methylated peptide (7 h of incubation) was digested with chymotrypsin resulting in the fragments 1-11 (GRGGFGGRGGF;

unmodified, 1024.53 Da) and 12-17 (RGGRGG; unmodified, 558.32 Da), each containing two arginine residues. Both fragments show the same methylation pattern with mainly two incorporated methyl groups (m/z 1052.5 and m/z 586.3), indicating an even distribution of the four methyl groups within the peptide (Fig. 4.3 B). To distinguish, whether the two arginines of the fragment 1-11 were either both monomethylated, or only one of them was dimethylated, or whether a mixture of both cases was present, the dimethylated fragment 1-11 (m/z 1052.5) was sequenced by MSMS (supplemental Fig. 4.2). Because of the repetitive amino acid sequence, the assignment of the peaks is not always clear as some internal fragments have the same mass like b-ions. The dominant peak at m/z 1007.5 (-45 Da of the parent ion) corresponds to the neutral loss of dimethylamin, indicating the asymmetric dimethylation of the peptide (17,18). Several ions could be uniquely assigned to either dimethylated Arg2 or Arg8. The internal ion GGRGG -17 and the ion b_7 +28 can only appear if Arg2 is dimethylated, whereas the ions a_5 -17, b_5, and b_7 were only possible with a dimethylated Arg8. There was no indication for a monomethylation of Arg2 or Arg8. Thus, it seems that PRMT8 generated mainly random pairs of dimethylated arginines within the peptide.

Methyltransferase activity of PRMT8 towards proteins in a cell lysate

All in all, the in vitro methyltransferase activity of recombinant PRMT8 towards the EWS

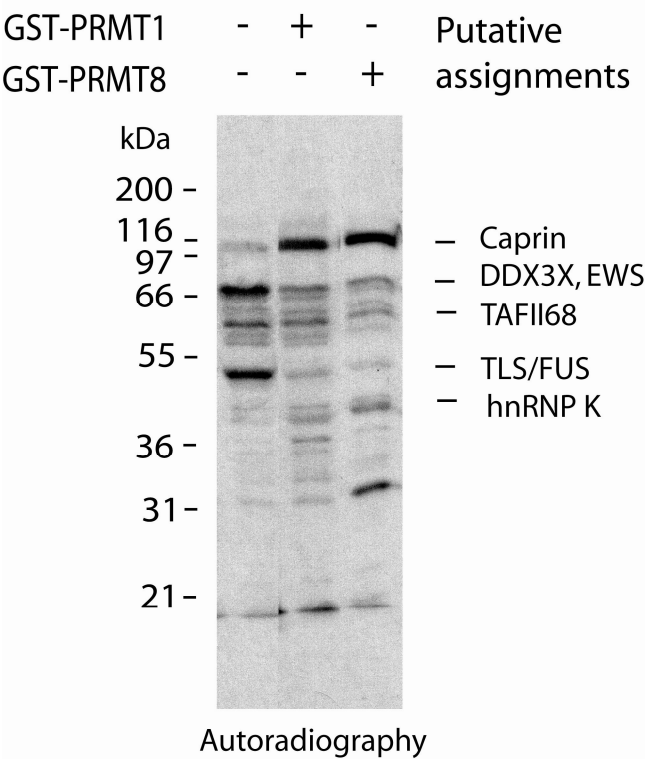


Figure 4.4: Methylation of a cell extract by PRMT1 and 8. Hypomethylated HEK cell lysate was incubated in the presence of ^{3}H-AdoMet without additional GST-PRMTs, with GST-PRMT1, or with GST-PRMT8 for 3 h at 30°C. Proteins were resolved by SDS-PAGE and incorporated labeled methyl groups were visualized by autoradiography after 28 h of exposure at -80°C. Molecular mass markers in kDa are indicated. The putative assignment of the proteins is based on the molecular mass of the identified interaction partners of PRMT8 (Table 1).

protein and the peptide was rather poor. To figure out, whether these are just bad substrates or recombinant GST-PRMT8 is indeed less active, total cell lysates from HEK cells were used as substrate for in vitro methylation in the presence of [^{3}H]-AdoMet. A total lysate from hypomethylated HEK cells was incubated with [^{3}H]-AdoMet and GST-PRMT8, GST-PRMT1, or without recombinant PRMTs. The proteins were separated by SDS-PAGE and proteins with incorporated methyl groups were visualized by autoradiography (Fig. 4.4). The endogenous methyltransferases in the hypomethylated cell extract were still active, as methylated proteins between 50 kDa and 66 kDa can be detected in the control (Fig. 4.4, lane 1). The addition of GST-PRMT1 or GST-PRMT8 to the hypomethylated cell extract yielded in methylation of several discrete proteins (Fig. 4.4, lanes 2 and 3). Both PRMT8 and PRMT1 strongly methylated a protein of 116 kDa. Between 55k Da and 80 kDa, the methylation pattern of PRMT1 and PRMT8 was comparable, although these bands were less intense than in the control. Among the smaller proteins, PRMT8 methylated two additional proteins of ~40 and 33 kDa, which are almost not methylated by PRMT1. The putative assignment (as indicated in Fig. 4.4) is based on the molecular mass of the identified, RGG-containing proteins from the pull-down experiments (Table 1).

PRMT8 interacts directly with the EWS protein independent of its methylation state

Although pull-down experiments indicated that unmethylated EWS protein interacted with PRMT8 (Fig. 4.1), in vitro methylation experiments revealed a rather poor methyltransferase activity of PRMT8 towards the EWS protein (Fig. 4.2). Therefore we investigated whether also completely methylated EWS protein, which is no more a substrate of PRMT8, still binds to PRMT8. Methylated EWS protein was obtained by expressing recombinant His-tagged EWS protein (see Experimental procedures) in HEK cells. MS analysis of the purified His-EWS construct revealed complete methylation of the EWS construct by the endogenous PRMTs of the HEK cells. By treating the cells with AdOx prior to transfection, unmethylated EWS protein was generated (proven by mass spectrometry). Mass shifts of unmethylated and completely methylated representative EWS peptides are shown in the mass spectra (Fig. 4.5 A). The pull-down with GST-PRMT8 was performed with such cell lysates from untreated or AdOx-treated HEK cells expressing the EWS construct. Binding of EWS protein was detected by western blotting using an EWS specific antibody (Fig. 4.5 B). Both, the unmethylated and the methylated His-EWS protein (110 kDa) were extracted by PRMT8 whereas no EWS protein was detected in the GST controls.

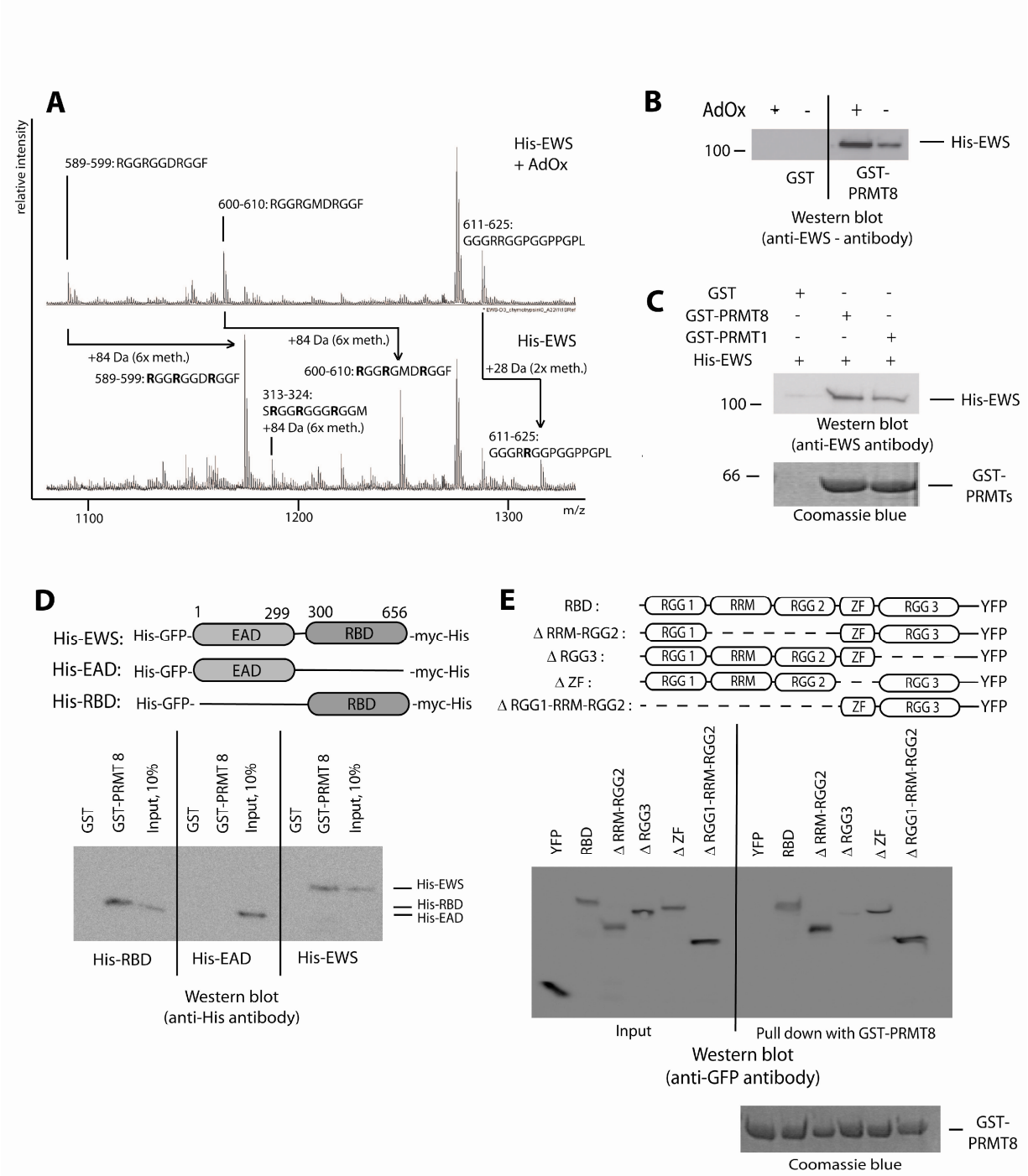


Figure 4.5: Identification of the interaction site of the EWS protein with PRMT8. **(A)** Cut-out of MALDI-TOF-MS spectra of the recombinant EWS protein expressed in HEK cells (lower spectrum) or HEK cells pre-treated with AdOx to inhibit endogenous methyltransferases (upper spectrum). The arrows indicate the mass shifts corresponding to one (+28 Da) or three (+84 Da) dimethylated arginines (labeled in bold (R)). **(B)** HEK cells, pre-treated or not with the methylation inhibitor AdOx, were transfected with His-GFP-EWS-myc-His construct, and the total lysate was used for a pull-down with GST-PRMT8 as bait. Bound recombinant EWS protein was detected by western blotting using an EWS specific antibody. **(C)** Purified His-GFP-EWS-myc-His was used for a pull-down with GST-PRMT8 as bait. Bound recombinant EWS protein was detected by western blotting using an EWS specific antibody. Similar amounts of GST-PRMTs were used as shown by coomassie blue staining. **(D)** Total HEK cell lysates, prior transfected with His-GFP-EWS-myc-His or truncated EWS constructs, was used for a pull-down with GST-PRMT8 as bait. Bound recombinant EWS constructs were detected by western blotting using a His-tag specific antibody.

PRMT1, a known interaction partner of the EWS protein (13), forms heterodimers with PRMT8 (4) and was identified in our pull-down experimets with PRMT8 (Table 1). So the interaction between PRMT8 and the EWS protein could be mediated via endogenous PRMT1 from the cell lysate. Thus, the pull-down experiments were performed by adding the baits GST-PRMT8 or GST-PRMT1, pre-bound to GSH-coupled agarose, to purified His-tagged EWS protein. PRMT-bound EWS protein was detected by subjecting the eluted proteins to SDS-PAGE followed by western blot analysis (Fig 4.5 C). The His-EWS protein bound to GST-PRMT1, but also to GST-PRMT8, whereas no binding of the EWS protein was observed in the GST control, demonstrating that the EWS protein is a direct interaction partner of PRMT8.

The EWS protein interacts with PRMT8 via its RGG-box 3

To narrow down the region of interaction between the EWS protein and PRMT8, truncated His-tagged GFP-EWS constructs, namely the N-terminal transcriptional activation domain (EAD, AA 1-299) and the C-terminal RNA-binding domain (RBD, AA 300-656) were expressed in HEK cells. The similar transfection efficiency of the two constructs was controlled by fluorescence microscopy (data not shown). The total lysates were used for pull-down experiments with GST-PRMT8 as bait. Bound EWS constructs were detected by western blotting using a His-tag-specific antibody (Fig. 4.5 D). Both the full-length EWS protein and the RBD interacted with PRMT8 whereas no signal was observed with the EAD. None of the constructs bound to the GST control. The RBD consists of three different RGG boxes (RGG 1-3), an RNA-recognition motif (RRM), and a putative C_2C_2 zinc finger motif (Fig. 4.5 E). By expressing truncated RBD-YFP fusion constructs (14) and extracting them from the cell lysate with GST-PRMT8, the region of interaction within the RBD could be identified to be the RGG box 3, as its deletion prevented binding to PRMT8. The deletion of RGG box 1, RRM, RGG box 2, or the zinc finger motif did not influence the binding to PRMT8 (Fig. 4.5 E).

Figure 4.5 (continued): (E) HEK cells were transfected with RBD-YFP constructs as indicated. The total lysate of transfected cells was used for pull-down experiments with GST-PRMT8 as bait. Bound recombinant RBD-YFP constructs were detected by western blotting using a GFP specific antibody, which recognizes also YFP. Similar amounts of GST-PRMT8 were used for the pull-down experiments as shown by coomassie blue staining. Expression of the constructs was shown by analyzing the total lysates directly by western blotting (input).

identification of PRMT1 as a binding partner of PRMT8 confirms the occurrence of heterodimers and proves the specificity of the used method. This interaction was also found by using GST-PRMT1 as bait, which pulled down endogenous PRMT8 from a HEK cell lysate (data not shown), although PRMT8 was as yet thought to be expressed only in brain tissue (4).

Further identified interaction partners of PRMT8 were the heterogeneous nuclear ribonucleoproteins (hnRNPs) K, U, H3, and A3. HnRNPs function in pre-mRNA splicing and mRNA transport and are known to form large complexes in the nucleus (19-21). HnRNP K has RGG motifs within its sequence which are known to be methylated by PRMT1 (22,23). A large scale proteomic analysis of interaction partners of hnRNP K identified nuclear and plasma membrane proteins (24), suggesting a function of hnRNP K in kinase-dependent signal transduction. Similarly, hnRNP U, which can be methylated and phosphorylated, was found beside its main localization in the nucleus to be attached to the plasma membrane, where it is a substrate of ecto-protein kinase (25). Protein arginine methylation is antagonized by phosphorylation (26) and vice versa (22), indicating that methylation of proteins at the plasma membrane, e.g. by PRMT8, can influence its phosphorylation and therefore its role in kinase-dependent signaling.

Caprin (cytoplasmic activation/proliferation-associated protein 1) is a cytoplasmic phosphoprotein, which is essential for cellular proliferation (27,28). In the brain, it was identified in messenger ribonucleoprotein particles (29). Recently it has been shown, that Caprin occurs also in post-synaptic granules in dendrites of neurons, where the RGG-rich C-terminal region of Caprin selectively binds mRNA (30). Whether caprin is methylated within its RGG boxes is not known yet. Considering, that the methylation of the hypomethylated cell extract with PRMT8 shows an intensive signal at 116 kDa (Fig. 4.4), it could well be, that caprin is a substrate of PRMT8.

Other specific interactors of PRMT8 were the structural proteins actin and tubulin. Tubulin fibers are usually used to transport cargo within the cell. It has been shown in neurons, that RNA is transported to the dendrites via the microtubules. Thereby, the RNA is bound to granules containing among others the RNA helicases DDX1, DDX3X, the EWS protein, TLS/FUS, but also hnRNP U (31). All of these proteins were identified in this study as PRMT8-interacting proteins. Thus, these proteins might rather co-purify by being part of a complex like such RNA-transporting granules than interacting directly with PRMT8. The specificity and functionality of the interaction with the remaining identified proteins like heat

shock proteins, tetrahydrofolate synthase, elongation factor 1, nesprin, and carbamoyl-phosphatase synthase needs further elucidation.

The three proteins TLS/FUS, EWS, and $TAF_{II}68$, which all belong to the TET-family of RNA-binding proteins, interact also with PRMT8. These multifunctional nuclear proteins are transcriptional co-factors, are involved in splicing (for a review, see (8)), and seem to have a function in DNA pairing and recombination (9). They all have RGG-repeats, the preferred methylation-motif, within their RNA-binding domain and are therefore potential substrates of PRMT8. In neurons, TLS/FUS was shown to localize in granules in the cytoplasm (32), and both proteins, TLS/FUS and EWS, were identified in mRNA-transporting granules in dendrites (31) as mentioned above. Finally, all TET-family proteins interact with the cytoplasmic kinase v-Src (33). As PRMT8 is expressed in particular in brain tissue and localized at the plasma membrane (4), an interaction between the TET-family proteins and PRMT8 as found in the present study seems feasible. Our results show, that PRMT8 and the EWS protein interact directly with each other via the RGG box 3 of the EWS protein (Fig. 4.6 E), and the co-localization of these two proteins in HEK cells indicate physiological relevance of this interaction. The in vitro methylation experiments showed a rather poor methyltransferase activity of PRMT8 towards the EWS protein (Fig. 4.2), although other proteins of a hypomethylated cell extract could be methylated by recombinant PRMT8 as efficient as by recombinant PRMT1 (Fig. 4.4). It seems that the interaction between PRMT8 and the EWS protein is not an enzyme-substrate interaction, as the binding between them is still maintained when the EWS protein is methylated.

It is known, that arginine methylation can influence protein-protein interaction. Methylation of Sam68 prevents its binding to SH3 domains (3), or methylation of hnRNP K inhibits binding to the kinase c-Src (22). However, the assembly of the hnRNP complex consisting of more than 20 different hnRNPs, most of which are methylated, seem to be independent of arginine methylation (20). Here we show for the first time that not only the binding of methylated proteins among each other, but also the interaction between a methyltransferase and its potential substrate protein may be maintained, although the substrate is completely methylated. PRMT8 might therefore act not only as a methyltransferase at the plasma membrane. It was shown for several nuclear proteins like nucleolin (34), or the already described TET-family proteins, that they are also localized at the cell membrane. However, their function and the mechanism, by which they are associated to the membrane is still unknown. They have no membrane crossing helices, and post-translational modifications,

which facilitate the membrane localization, could not be identified so far for these proteins. PRMT8 might therefore act as an adapter protein, which target not only PRMT1 activity, as proposed by Lee et al. (4), but also other nuclear proteins to the membrane, although they are already methylated.

Acknowledgements

We are grateful to Doris Grossenbacher for her support in the production and purification of the recombinant proteins. We would like to thank Dr. Peter Gehrig, Dr. Bernd Roschitski, and Dr. Bertran Gerrits from the FGCZ for their contributions in recording and interpreting the MS data, and Dr. Philipp Christen for helpful discussions and critical reading of the manuscript. This work was supported by the Olga Mayenfisch Stiftung, Hartmann Müller-Stiftung für medizinische Forschung, and the Julius Müller Stiftung zur Unterstützung der Krebsforschung.

The abbreviations used are: GST, Glutathione-S-transferase; GAR, glycine and arginine rich region; MALDI, matrix-assisted laser desorption ionization; TOF, time of flight; MS, mass spectrometry; GFP, green fluorescent protein; YFP, yellow fluorescent protein; α-CCA, α-cyano-4-hydroxycinnamic acid.

References

1. Paik, W. K., Paik, D. C., and Kim, S. (2007) *Trends in biochemical sciences* 32(3), 146-152
2. Bedford, M. T., and Richard, S. (2005) *Molecular cell* 18(3), 263-272
3. Pahlich, S., Zakaryan, R. P., and Gehring, H. (2006) *Biochimica et biophysica acta* 1764(12), 1890-1903
4. Lee, J., Sayegh, J., Daniel, J., Clarke, S., and Bedford, M. T. (2005) *J Biol Chem* 280(38), 32890-32896
5. Belyanskaya, L. L., Gehrig, P. M., and Gehring, H. (2001) *Journal of Biological Chemistry* 276(22), 18681-18687
6. Thomas, G. R., Faulkes, D. J., Gascoyne, D., and Latchman, D. S. (2004) *Biochem Biophys Res Commun* 318(4), 1045-1051
7. Lee, J., Rhee, B. K., Bae, G. Y., Han, Y. M., and Kim, J. (2005) *Stem cells (Dayton, Ohio)* 23(6), 738-751
8. Law, W. J., Cann, K. L., and Hicks, G. G. (2006) *Briefings in functional genomics & proteomics* 5(1), 8-14
9. Li, H., Watford, W., Li, C., Parmelee, A., Bryant, M. A., Deng, C., O'Shea, J., and Lee, S. B. (2007) *J Clin Invest*
10. Belyanskaya, L. L., Delattre, O., and Gehring, H. (2003) *Experimental cell research* 288(2), 374-381
11. May, W. A., Gishizky, M. L., Lessnick, S. L., Lunsford, L. B., Lewis, B. C., Delattre, O., Zucman, J., Thomas, G., and Denny, C. T. (1993) *Proceedings of the National Academy of Sciences of the United States of America* 90(12), 5752-5756
12. Alex, D., and Lee, K. A. (2005) *Nucleic Acids Res* 33(4), 1323-1331
13. Pahlich, S., Bschir, K., Chiavi, C., Belyanskaya, L., and Gehring, H. (2005) *Proteins* 61(1), 164-175

14. Zakaryan, R. P., and Gehring, H. (2006) *Journal of molecular biology* 363(1), 27-38
15. Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. (1999) *Electrophoresis* 20(18), 3551-3567
16. Najbauer, J., and Aswad, D. W. (1990) *J Biol Chem* 265(21), 12717-12721
17. Gehrig, P. M., Hunziker, P. E., Zahariev, S., and Pongor, S. (2004) *Journal of the American Society for Mass Spectrometry* 15(2), 142-149
18. Rappsilber, J., Friesen, W. J., Paushkin, S., Dreyfuss, G., and Mann, M. (2003) *Anal Chem* 75(13), 3107-3114
19. Dreyfuss, G., Matunis, M. J., Pinol-Roma, S., and Burd, C. G. (1993) *Annual review of biochemistry* 62, 289-321
20. Pawlak, M. R., Banik-Maiti, S., Pietenpol, J. A., and Ruley, H. E. (2002) *J Cell Biochem* 87(4), 394-407
21. Krecic, A. M., and Swanson, M. S. (1999) *Current opinion in cell biology* 11(3), 363-371
22. Ostareck-Lederer, A., Ostareck, D. H., Rucknagel, K. P., Schierhorn, A., Moritz, B., Huttelmaier, S., Flach, N., Handoko, L., and Wahle, E. (2006) *J Biol Chem* 281(16), 11115-11125
23. Chiou, Y. Y., Lin, W. J., Fu, S. L., and Lin, C. H. (2007) *The protein journal* 26(2), 87-93
24. Mikula, M., Dzwonek, A., Karczmarski, J., Rubel, T., Dadlez, M., Wyrwicz, L. S., Bomsztyk, K., and Ostrowski, J. (2006) *Proteomics* 6(8), 2395-2406
25. Jordan, P., Heid, H., Kinzel, V., and Kubler, D. (1994) *Biochemistry* 33(49), 14696-14706
26. Hsu Ia, W., Hsu, M., Li, C., Chuang, T. W., Lin, R. I., and Tarn, W. Y. (2005) *J Biol Chem* 280(41), 34507-34512
27. Wang, B., David, M. D., and Schrader, J. W. (2005) *J Immunol* 175(7), 4274-4282
28. Grill, B., Wilson, G. M., Zhang, K. X., Wang, B., Doyonnas, R., Quadroni, M., and Schrader, J. W. (2004) *J Immunol* 172(4), 2389-2400
29. Angenstein, F., Evans, A. M., Ling, S. C., Settlage, R. E., Ficarro, S., Carrero-Martinez, F. A., Shabanowitz, J., Hunt, D. F., and Greenough, W. T. (2005) *J Biol Chem* 280(8), 6496-6503
30. Solomon, S., Xu, Y., Wang, B., David, M. D., Schubert, P., Kennedy, D., and Schrader, J. W. (2007) *Molecular and cellular biology* 27(6), 2324-2342
31. Kanai, Y., Dohmae, N., and Hirokawa, N. (2004) *Neuron* 43(4), 513-525
32. Belly, A., Moreau-Gachelin, F., Sadoul, R., and Goldberg, Y. (2005) *Neuroscience letters* 379(3), 152-157
33. Lee, H. J., Kim, S., Pelletier, J., and Kim, J. (2004) *FEBS letters* 564(1-2), 188-198
34. Hovanessian, A. G., Puvion-Dutilleul, F., Nisole, S., Svab, J., Perret, E., Deng, J. S., and Krust, B. (2000) *Experimental cell research* 261(2), 312-328

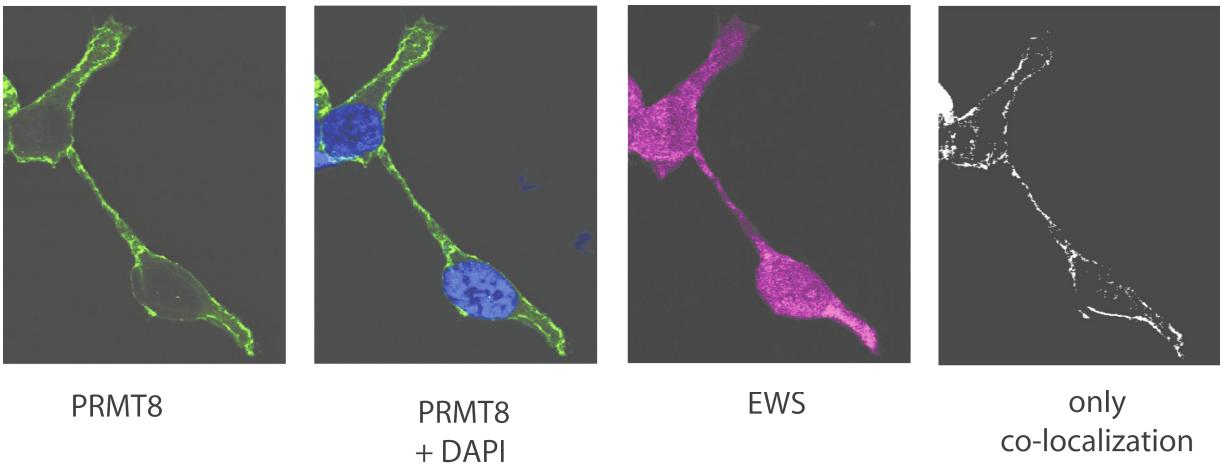


Figure 4.6: Co-localization of PRMT8-GFP and endogenous EWS protein. HEK cells were transiently transfected with PRMT8-GFP. Endogenous EWS protein is visualized by immunofluorescence using a rabbit anti-EWS (C-terminus) primary antibody and indocarbocyanine (Cy3)-coupled goat anti-rabbit secondary antibodies. The fluorescent proteins were analyzed by confocal microscopy using excitation-emission wavelengths of 405-470 nm, 488-509 nm and 550-570 nm for DAPI, GFP and Cy3, respectively.

Co-localization of PRMT8 and the EWS protein

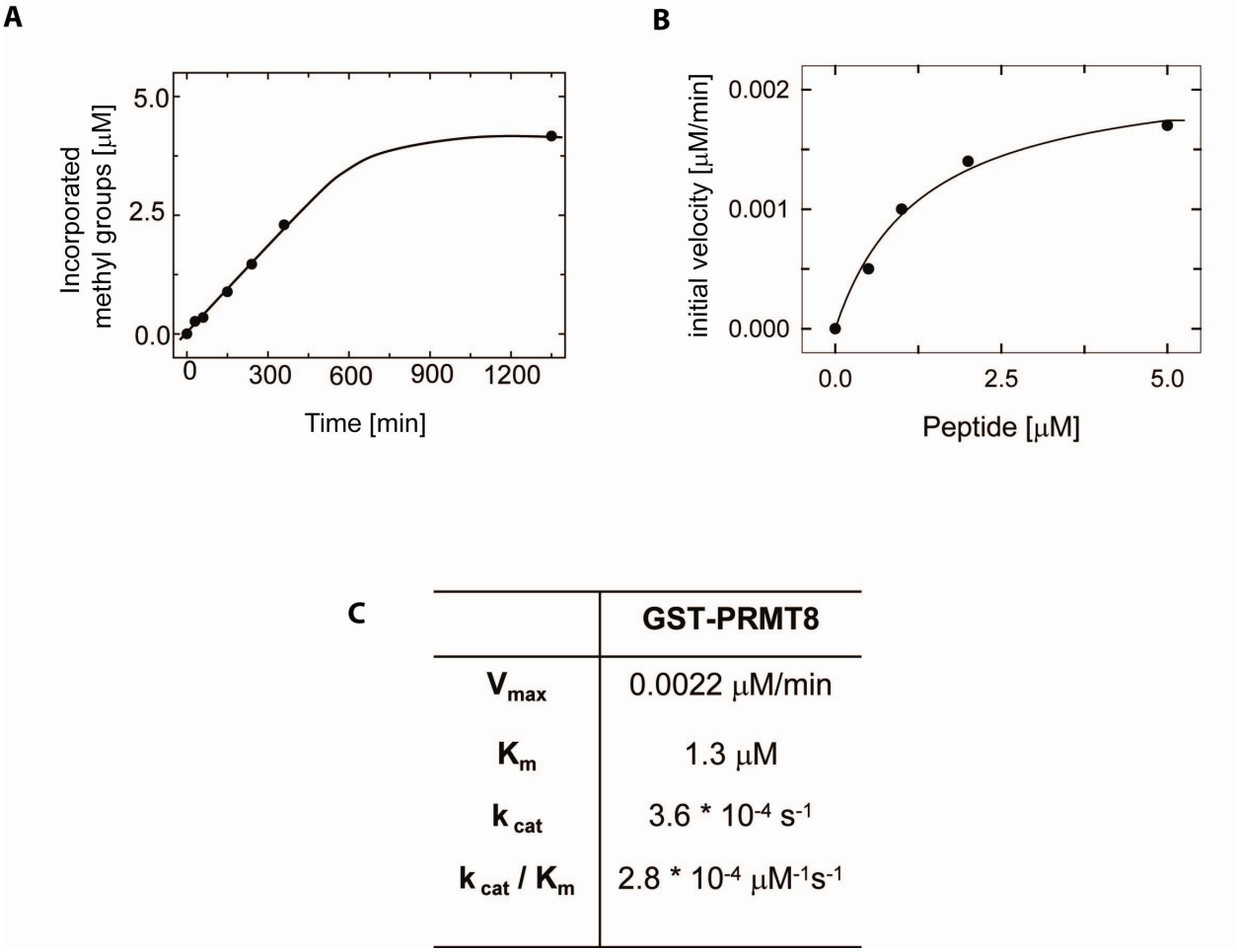
To visualize the interaction of PRMT8 and the EWS protein, HEK cells were transfected with GFP-PRMT8, which shows the typical membrane-associated localization (4) (Fig. 4.6). Endogenous EWS protein was visualized within the same cells by immunofluorescence using an antibody which recognizes the C-terminus of the EWS protein. Both nuclear and extranuclear EWS protein was detected. The merged pictures indicate, that large parts of PRMT8 co-localize with the EWS protein at the plasma membrane, supporting the notion of a functional intracellular interaction between the EWS protein and PRMT8.

Discussion

PRMT8 is the only membrane-bound member of the human PRMT family (4), but neither endogenous substrate proteins nor its particular function at the cell membrane is known so far. In the present study, we identified new binding partners and therefore potential substrates of PRMT8 by GST pull-down experiments.

More than 20 different proteins which were pulled down with GST-PRMT8 could be identified by mass spectrometry (Table 1). Among them was the arginine methyltransferase PRMT1. PRMT8 and PRMT1 have been reported to form homo- and heterodimers (4). Our

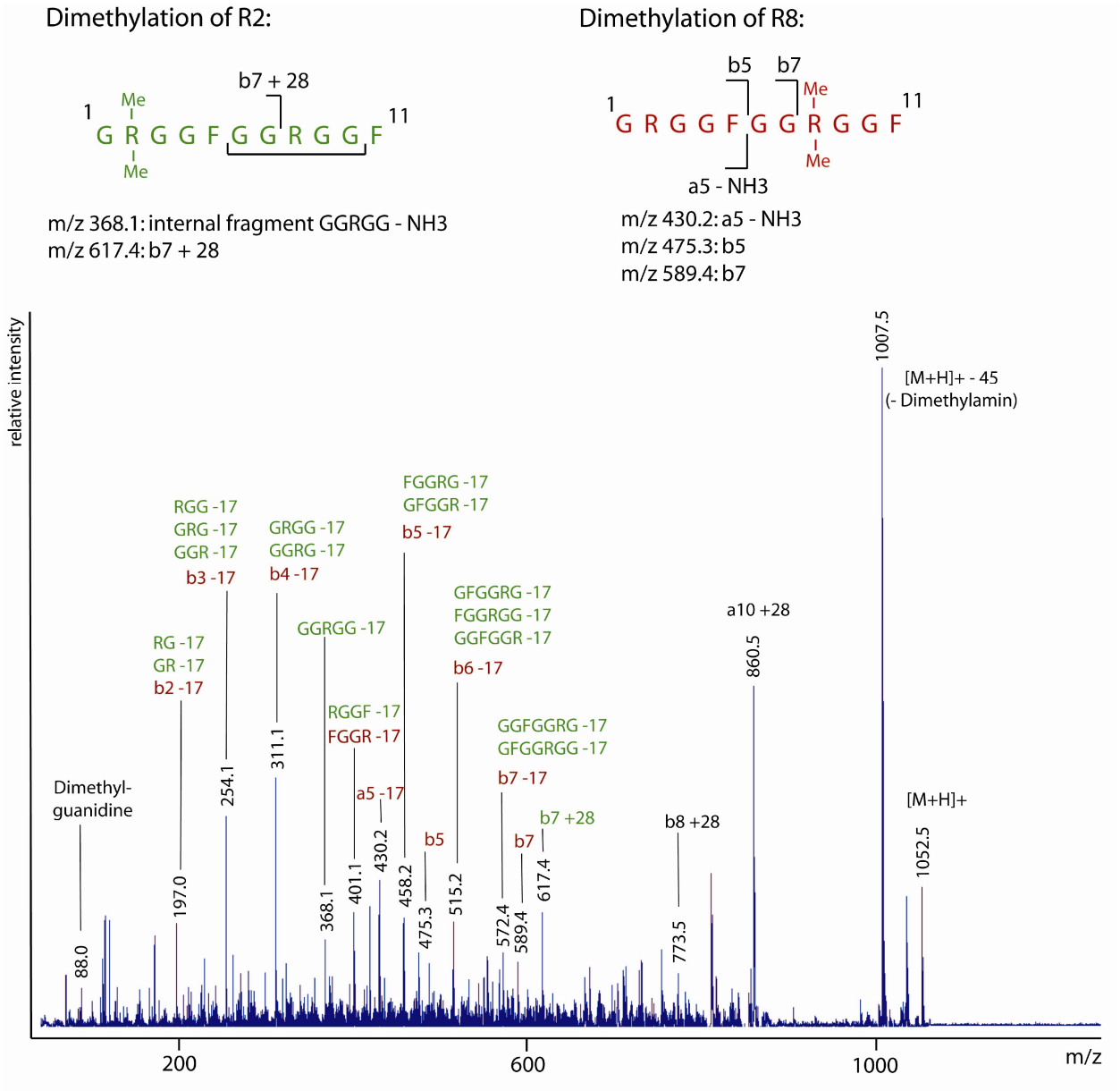
Supplemental Data



C

	GST-PRMT8
V_{max}	0.0022 μM/min
K_m	1.3 μM
k_{cat}	$3.6 * 10^{-4}$ s^{-1}
k_{cat} / K_m	$2.8 * 10^{-4}$ $\mu M^{-1}s^{-1}$

Supplemental Figure 4.1: Kinetics of the peptide methylation by PRMT8. **(A)** The incorporated methyl groups were quantified from the MALDI-TOF-MS spectra at each time point (Fig. 4.3 A) as described (13) resulting in a linear initial phase of the peptide methylation. The slope of the linear initial phase corresponds to the initial velocity. Complete methylation of the peptide would correspond to 8 μM incorporated methyl groups. **(B)** The initial velocity of the peptide methylation was determined using 0.1 μM PRMT8 and varying concentrations of peptide as indicated. **(C)** Kinetic parameters were calculated by hyperbolic curve fitting (SigmaPlot) of (B) as described (13).



Supplemental Figure 4.2: MSMS-spectrum of the dimethylated peptide fragment 1-11 at m/z 1052.5. The two methyl groups in this fragment can be situated either at arginine 2 or at arginine 8. Ions and internal fragments which can be assigned to the dimethylated Arg2 are labeled in green, those which can be assigned to the dimethylated Arg8 are labeled in red. Ions which fit to both cases are labeled in black.

5. Analysis of Ewing Sarcoma (EWS)-binding Proteins: Interaction with hnRNP M, U, and RNA-helicases p68/72 within Protein-RNA complexes

Steffen Pahlich, Lilan Quero, Rouzanna P Zakaryan, and Heinz Gehring

In preparation for publication

Summary

The human Ewing Sarcoma (EWS) protein belongs to the TET family of RNA-binding proteins and consists of an N-terminal transcriptional activation domain (EAD) and a C-terminal RNA-binding domain (RBD), which is extensively methylated at arginine residues. This multifunctional protein acts in transcriptional co-activation, DNA-recombination, -pairing and -repair, but also a role in splicing and mRNA transport was reported. The role of arginine methylation within these functions is still unclear.

In this study, we performed pull-down experiments with methylated and unmethylated recombinant EWS protein as bait, and identified the interacting proteins by mass spectrometry. Methylated EWS protein was obtained by co-expression of protein arginine methyltransferase (PRMT) 1 and the EWS protein in E.coli. Both, unmethylated and methylated EWS protein interact via its C-terminal RNA-binding domain (RBD) with protein-RNA complexes consisting of mainly heterogeneous nuclear ribonucleoproteins (hnRNPs) and RNA helicases. HnRNP M and U, the RNA-helicases p68 (DDX5) and p72 (DDX17), but also actin and tubulin seem to interact directly with the EWS protein.

The interaction of the EWS protein with p68 was confirmed by co-precipitation of the two proteins and the RBD of the EWS protein was identified to be the interacting domain. Co-localization of the EWS protein and the RNA-helicases in the nucleus was visualized by co-expression of fluorescently labeled EWS protein and p68 or p72 in HEK 293 (T) cells. Overexpression of the RNA-helicases relocated the fluorescent EWS protein from the nucleoplasm to nucleolar capping structures, a relocalization observed for many proteins, when transcription is inhibited. Remarkably, extensive arginine methylations within the RBD of the EWS protein, which occur short after or during translation, does not change the interaction pattern.

Introduction

The human Ewings sarcoma (EWS) protein belongs to the TET-family of RNA-binding proteins. It consists of an N-terminal transcriptional activation domain (EAD) and a C-terminal RNA-binding domain. The EAD interacts with calmodulin (42), the splicing factor 1 (43), protein kinases (42,44), and RNA polymerase II (1), and function as a transcriptional co-activator (2-4). In Ewings sarcoma, a tumor-associated chromosomal translocation of the *EWS* gene results in the expression of fusion proteins of the EWS EAD with DNA-binding domains of several ETS transcription factors with transforming potential (5,6). The C-terminal RNA-binding domain (RBD) of the wild type EWS protein, which represses the tumor-associated transcriptional activity of the EAD (7), consists of a RNA-recognition motif (RRM), a putative C_2C_2 zinc finger, and three arginine-glycine rich domains (RGG boxes 1-3). Additionally to the transcriptional activity of the EWS protein, a role in splicing and mRNA transport is proposed, as the EWS protein interacts with the splicing factors TASR-1 and -2 (8,9) and YB-1 (10), it co-purifies with the hnRNPs A1 and C1/C2 (11), and it was identified in a large scale proteomic analysis of the human spliceosome (12). Recently, a role in DNA pairing (13) and recombination was described, as EWS protein-deficient mice fail to develop functional B-lymphocytes and have defects in meiosis (14).

The EWS protein is located mainly in the nucleus, directed by its C-terminal nuclear localization sequence (15) via transportin (16). Minor fractions of the protein were also identified in the cytosol and at the cell membrane. Independent on its subcellular localization, the RGG boxes of the EWS protein were found to contain 30 asymmetrically dimethylated arginines (17). Arginine methylation is a eukaryotic post-translational modification catalyzed by a family of enzymes called protein arginine methyltransferases (PRMTs). They catalyze the transfer of a methyl group from the donor S-adenosyl-L-methionine to the guanidino nitrogens of the arginine side chain. Methylated arginines are found mainly in RGG-motifs of RNA-binding proteins like heterogeneous nuclear ribonucleoproteins (hnRNPs), RNA helicases, spliceosomal proteins, and histones. Methylated proteins function in transcriptional regulation, pre-m-RNA splicing and mRNA-transport, translation, but also a role in signaling is proposed (for reviews, see (18,19)). However, for many proteins, like the EWS protein, the role of methylation is still not known.

In this study, we performed pull-down experiments with unmethylated and methylated recombinant EWS protein as bait to identify interaction partner in dependence of this modification. Both, unmethylated and methylated EWS protein interacts with large protein

complexes via its RBD. The interaction with these RNAse-sensitive complexes, consisting of mainly RNA-binding proteins, could be mediated via RNA-binding of the EWS protein, or via the interaction of the EWS protein with hnRNP M, U, and the RNA-helicases p68 / p72. Co-precipitation and co-localization studies of the EWS protein and p68 did not only confirm the interaction, but showed also a relocalization of the EWS protein upon p68 expression from the nucleoplasm to the nucleolar periphery. We could show that the EWS protein is methylated immediately after or even co-transcriptionally in the cell. The extensive arginine methylations within the RBD of the EWS protein are neither needed for subcellular localization (15) nor for protein-protein interaction, as demonstrated in this study. Therefore, a role in RNA-binding, in affecting the activation/repression activity, or even in stabilization of the EWS protein is proposed.

Materials and Methods

Plasmids and Constructs –The cDNA of EWS was kindly provided by Dr. Olivier Delattre (Institut Curie, Pathologie Moléculaire des Cancers, Paris Cedex). GST-p68 and GST-p72 expression vectors (pEBG2T) for expression in mammalian cells were a generous gift of Dr. Frances Fuller-Pace (Division of Pathology & Neuroscience, University of Dundee, UK), and His-tagged p68 expression vector (pET-30a) was kindly obtained from Dr. Zhi-Ren Liu (Department of Animal and Dairy Sciences, Auburn University, Alabama).

Expression of recombinant proteins – GST-EWS was expressed in E.coli as described (21). Co-expression of GST-EWS (expression vector pGEX-6p-2) and His-PRMT1 (expression vector pET-28c) was done in E.coli XL-blue cells. His-tagged p68 was expressed in E.coli BL1-codon plus strain. For expression of His-EWS constructs (15), Human embryonic kidney (HEK) 293 T cells were transiently transfected using the calcium phosphate precipitation method. Efficient transfection and proper subcellular localization of the different constructs was checked by fluorescence microscopy.

Cell culture, preparation of total cell lysates – Jurkat cells were cultivated in RPMI medium at 37°C. Cells were harvested, washed twice in ice cold PBS buffer and lysed for 30 min on ice in lysis buffer A (50 mM Tris, pH 8.0; 200 mM NaCl; 1 mM PMSF; 1 mM DTT; 1 % Triton-X 100; protease inhibitor mix (Roche)). The lysate was centrifuged for 30 min at 13000 rpm and the supernatant, designated as total Jurkat lysate, was used for pull-down experiments. HEK cells were cultivated in Dulbecco's Modified Eagle's Medium (DMEM)

supplemented with 10% (v/v) fetal calf serum (FCS) in a humified 10% (v/v) CO_2 atmosphere at 37 °C. The cells were cultured to 40% confluency on glass cover slips pre-treated with poly-L-Lysin (30 µg/ml) prior to transfection.

Immunostaining and confocal microscopy - 24 h after the transfection, HEK cells were washed with PBS, fixed with 4% paraformaldehyd, permeabilized with 1% Triton X-100 and incubated for 15 min at RT. The cells were washed three times with PBS, non-specific binding sites were blocked in 10% FCS, 0.1% Glycin in PBS, pH 7.4 for one h and washed again three times with PBS. Immunostaining was performed using a primary goat anti-GST antibody (amersham, 1:2000) and a secondary Cy3-coupled anti-goat antibody (Jackson Immuno Research, 1:400). DNA was stained with 4',6-diamidino-2-phenylindol (DAPI, Roche). Laser-scanning confocal fluorescence microscopy was performed as described (15) using excitation-emission wavelength of 405-470 nm for DAPI, 514-528 nm for YFP and 543-570 nm for Cy3. The pictures were merged and co-localized with Adobe Photoshop 8.0 and Imaris 5.7.0, respectively.

GST pull-down - GST or GST-EWS (10 µg) was bound to glutathione-coupled agarose (Sigma) at RT (2 h) and afterwards, the agarose was washed twice with 100mM Na-Borate and twice with 50mM Tris, pH 8.0. Total Jurkat cell lysate (1 mg of total protein) was added in the presence of binding buffer (50 mM Tris, pH 8.0; 100 mM NaCl; 1 mM PMSF; 1 mM DTT; 0.1 % Triton X-100; protease inhibitor mix (Roche)). If indicated, RNase A was added to a final concentration of 0.2 mg/ml. The sample was incubated for one h at RT and one h at 4°C under constant shaking. The agarose was washed three times with binding buffer and bound proteins were specifically eluted with 20 mM glutathione / 50 mM Tris, pH 8.0. The eluted proteins were separated by SDS-PAGE (10% acrylamide) and analyzed by mass spectrometry after in-gel digest or by western blotting.

Mass spectrometry (MS) - MS analyses were performed at the Functional Genomics Center Zürich (FGCZ) and at the Protein Analysis Unit of the University of Zürich as described (21).

Briefly, for identification of EWS-binding proteins from the GST pull-down, the whole gel lane was cut into slices, the proteins were digested with trypsin, the pooled peptides were desalted using C18-ZipTips (Millipore, Bedford, MA), and analyzed by nano-LC-MSMS using an LTQ-FT (Thermo Electron, Bremen, Germany) as described (Pahlich et al. submitted for publication). MALDI-TOF-MS analysis was performed on a Bruker-Daltonics Autoflex II

MALDI mass spectrometer operated in the positive ion reflector mode using α-cyano-hydroxycinnamic acid (4mg/ml in 70% acetonitrile/0.1% TFA) as matrix.

Results

GST pull-down with EWS protein as bait

To identify endogenous interaction partner of the EWS protein, GST pull-down experiments were performed. Recombinant GST-EWS protein, expressed in E.coli, was bound to proteins were specifically eluted with glutathione and subjected to SDS-PAGE followed by coomassie blue staining of the gel (Fig. 5.1, lane 1). Beside the GST-EWS protein band at ~116 kDa, several co-purifying protein bands can be seen which are not present when GST was used as bait (lane 3). To identify the co-purified proteins, the whole gel lane was cut into slices, each slice was in-gel digested, and the resulting peptides were analyzed by LC-MSMS as described in materials and methods. More than 30 different proteins were identified, among them mainly RNA-binding proteins like hnRNPs or RNA-helicases (Table 1), but also the methyltransferase PRMT1 and the structural proteins actin, tubulin, and vimentin. All identified proteins including their Mascot scores are listed in supplemental table 1. Among the identified proteins, the hnRNPs A0, A1, D0, G, and H, but also TLS/FUS and $TAF_{II}68$ were found to contain methylated arginine residues (Table 2). The identified methylation sites of hnRNP H (R232) and $TAF_{II}68$ (R475, R483, R528, R535, R570) have not been described to our knowledge, whereas the other identified methylation sites have been reported (20).

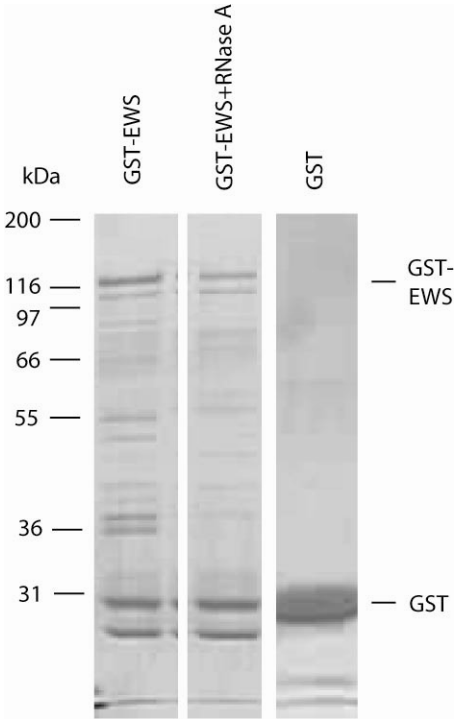


Figure 5.1: Pull-down with GST-EWS protein as bait. Total cell lysate from Jurkat cells was used for a pull-down experiment with recombinant GST-EWS protein in the absence (lane 1) or presence (lane 2) of RNase A, or recombinant GST (lane 3) as bait. Extracted proteins were resolved by SDS-PAGE and visualized by Coomassie blue staining of the gel.

Tab. 1: List of proteins identified in the indicated pull-down experiments. Proteins identified in at least two independent experiments are labeled in bold (**X**), proteins identified in only one experiment are labeled in roman (X).

Accession number	Protein description	GST-EWS	GST-EWS + RNase A	methylated GST-EWS	methylated GST-EWS + RNase A
ANM1_HUMAN	(Q99873) Protein arginine N-methyltransferase 1	**X**	**X**	**X**	**X**
ROA0_HUMAN	(Q13151) Heterogeneous nuclear ribonucleoprotein A0	**X**		X	
ROA1_HUMAN	(P09651) Heterogeneous nuclear ribonucleoprotein A1	**X**		**X**	
ROA2_HUMAN	(P22626) Heterogeneous nuclear ribonucleoproteins A2/B1	**X**		**X**	
ROA3_HUMAN	(P51991) Heterogeneous nuclear ribonucleoprotein A3	**X**		**X**	
ROAA_HUMAN	(Q99729) Heterogeneous nuclear ribonucleoprotein A/B	**X**		X	
HNRPD_HUMAN	(Q14103) Heterogeneous nuclear ribonucleoprotein D0	**X**		**X**	
HNRPF_HUMAN	(P52597) Heterogeneous nuclear ribonucleoprotein F	**X**		**X**	
HNRPG_HUMAN	(P38159) Heterogeneous nuclear ribonucleoprotein G	**X**		**X**	
HNRH1_HUMAN	(P31943) Heterogeneous nuclear ribonucleoprotein H	**X**		**X**	
HNRH2_HUMAN	(P55795) Heterogeneous nuclear ribonucleoprotein H'	**X**		X	
HNRH3_HUMAN	(P31942) Heterogeneous nuclear ribonucleoprotein H3	**X**		X	
HNRPM_HUMAN	(P52272) Heterogeneous nuclear ribonucleoprotein M	**X**	**X**	**X**	**X**
HNRPQ_HUMAN	(O60506) Heterogeneous nuclear ribonucleoprotein Q	**X**		X	
HNRPU_HUMAN	(Q00839) Heterogeneous nuclear ribonucleoprotein U	**X**	X	**X**	**X**
CIRBP_HUMAN	(Q14011) Cold-inducible RNA-binding protein (hnRNP A18)	**X**			
FUS_HUMAN	(P35637) RNA-binding protein FUS (hnRNP P)	**X**	**X**	**X**	
DDX5_HUMAN	(P17844) ATP-dependent RNA helicase p68 (DDX5)	**X**	**X**	**X**	**X**
DDX17_HUMAN	(Q92841) ATP-dependent RNA helicase p72 (DDX17)	**X**	**X**	**X**	X
DDX3X_HUMAN	(O00571) ATP-dependent RNA helicase DDX3X	**X**	**X**		
DHX9_HUMAN	(Q08211) ATP-dependent RNA helicase A	**X**		**X**	X
Q8IYE5	(Q8IYE5) ATP-dependent RNA helicase DHX36	**X**			
ACTB_HUMAN	(P60709) Actin, cytoplasmic 1 (Beta-actin)	**X**	**X**	**X**	**X**
TBAK_HUMAN	(P68363) Tubulin alpha-ubiquitous chain	**X**		**X**	
TBB2_HUMAN	(P07437) Tubulin beta-2 chain	**X**	**X**	**X**	**X**
VIME_HUMAN	(P08670) Vimentin	**X**		**X**	

RNase-sensitivity of the complex formation with the EWS proteins

HnRNPs are known to form protein-RNA complexes in the nucleus. To test, whether the identified proteins bind directly to the EWS protein or RNA-protein complexes are co-purified, the cell lysate was pre-treated with RNase A. Subjecting the co-purified proteins to SDS-PAGE displays the absence of several proteins (Fig. 5.1, lane 2). MS analysis after in-gel digestion revealed that almost all RNA-binding proteins except hnRNP M, hnRNP U, and the RNA-helicases p68, p72 were absent after RNase treatment (Table 1, supplemental table 1).

The EWS protein consists of an N -terminal EAD and a C-terminal RBD. To determine, which domain is interacting with the RNA-protein complex, pull-down experiments with

Table 2: Methylated peptides of proteins identified in the EWS pull-down. The methylation sites of hnRNP H and TAF(II)68 are not described so far.

Accession number	Protein	m/z	Peptide sequence	Modification	
Q13151	hnRNP A0	1968.85	SNSGPY**R**GGYGGGGGYGGSSF	1 dimethylation (R291)	
P09651	hnRNP A1	1422.67	QEMASASSSQ**R**GR	1 dimethylation (R193)	
		2058.92	SGSGNFGGG**R**GGGFGGNDNFGR	1 dimethylation (R205)	
		1340.64	GGNFSG**R**GGFGGSR	1 dimethylation (R224)	
Q14103	hnRNP D0	1334.65	**R**GGHQNSYKPY	1 dimethylation (R345)	
P38159	hnRNP G	770.47	APVS**R**GR	1 dimethylation (R185)	
P31943	hnRNP H	2901.24	**R**GAYGGGYGGYDDYNGYNDGYGFGSDR	1 dimethylation (R232)	new
P35637	TLS/FUS	1762.85	GG**R**G**R**GGSGGGGGGGGGGYNR	2 dimethylations (R216, R218)	
Q92804	TAF(II)68	2189.91	GGGYGGD**R**GGGYGGD**R**GGYGGDR	2 dimethylations (R475, R483)	new
		1428.62	GGGYGGD**R**GGYGGDR	1 dimethylation (R475)	new
		1843.85	S**R**GGYGGD**R**GGGSGYGGDR	2 dimethylations (R528, R535)	new
		1285.59	GGGYGGD**R**GGYGGK	1 dimethylation (R570)	new

individual domains in the presence and absence of RNase were performed. The same set of hnRNPs and RNA-helicases were identified with the RBD as bait, whereas none of them bound to the EAD (data not shown).

GST pull-down with methylated EWS protein as bait

The RBD consists of an RNA-recognition motif (RRM), a putative zinc finger, and three RGG rich domains (RGG-boxes 1-3), which contain 30 asymmetrically dimethylated arginines (17,21). To investigate the influence of arginine methylation of the EWS protein on binding its partners, methylated recombinant EWS protein was used as bait. It was generated by co-expression of GST-EWS and His-PRMT1 in E.coli. The GST-EWS protein was affinity-purified and its methylation state was controlled by MALDI-TOF-MS. All peptides in the spectrum, which contained potential arginine methylation sites were found to be methylated (Fig. 5.2). Although some incomplete methylated peptides were present, their complete methylation was dominant. The EWS protein co-expressed with PRMT1, designated here as methylated EWS protein was used for pull-down experiments as described above. A very similar set of proteins was extracted and identified when methylated EWS protein was used as bait in the presence or absence of RNase A (Table 1, supplemental table 1).

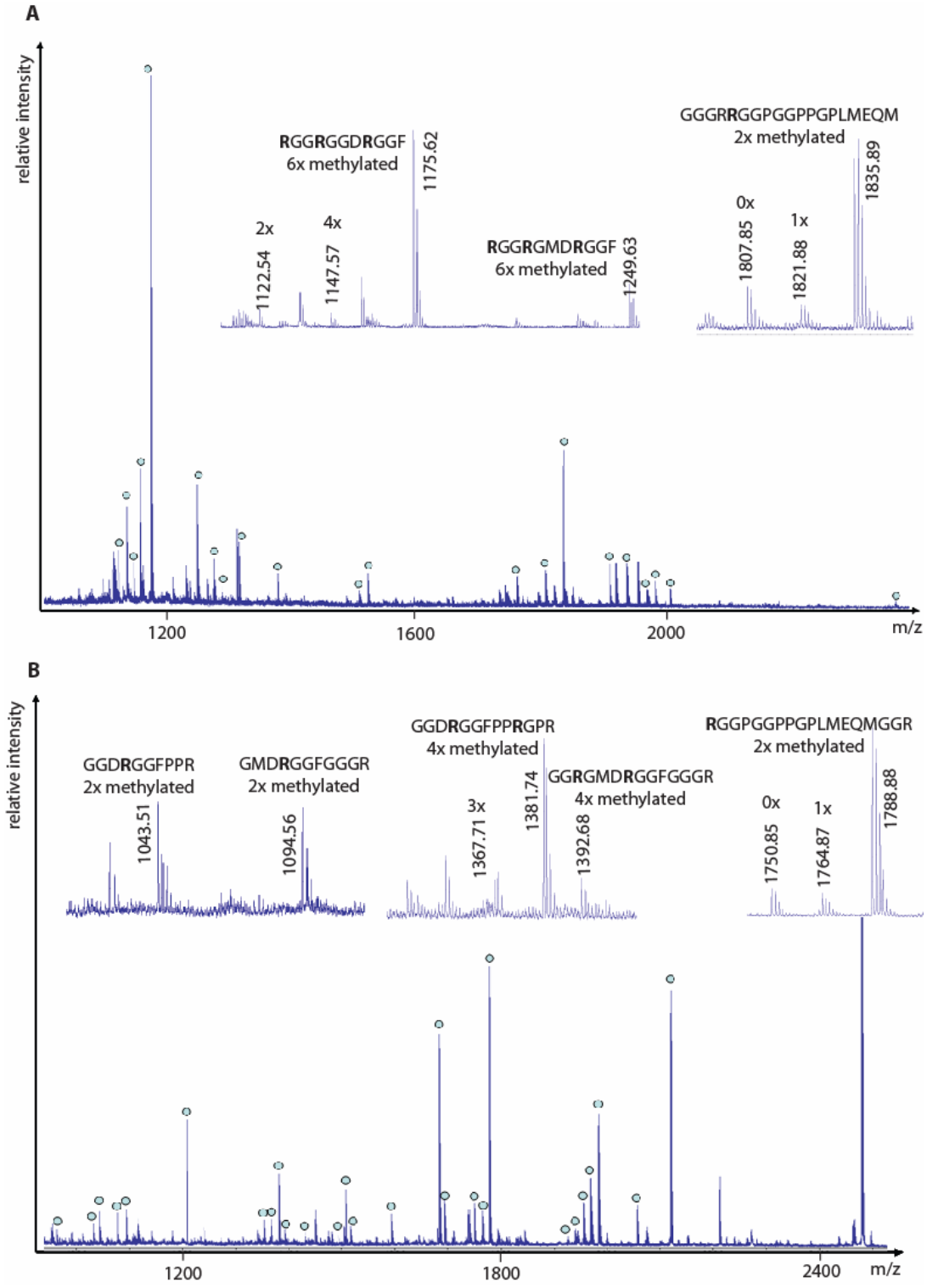


Figure 5.2: Methylation state of recombinant EWS protein co-expressed with His-PRMT1. GST-EWS protein was co-expressed with His-PRMT1 in E.coli, the EWS protein was affinity-purified using glutathione-coupled agarose, the extracted EWS protein, resolved by SDS-PAGE, was in gel digested either with trypsin or chymotrypsin, and the resulting peptides were subjected to MALDI-TOF-MS analysis. (A) Mass spectrum of the tryptic peptides. All peaks that could be assigned to the GST-EWS protein are labeled with a circle. Peptides with potential methylation sites (**R**) and the corresponding methylation state are enlarged. (B) Mass spectrum of the chymotryptic peptides. All peaks that could be assigned to the GST-EWS protein are labeled with a circle. Peptides with potential methylation sites (**R**) and the corresponding methylation state are enlarged.

Interaction of the EWS protein with RNA helicase p68

In the pull-down experiments, the RNA helicase p68 was found to interact with the RBD of the EWS protein independent on its methylation or RNase A treatment. To confirm this interaction, pull-down experiments with affinity-purified recombinant proteins were performed. Purified recombinant GST-RBD was incubated with purified recombinant His-tagged p68 and bound p68 was detected by western blotting using anti-His antibodies (Fig. 5.3). Binding of the p68 was observed with both, unmethylated and methylated RBD as bait, whereas no p68 could be detected in the GST control. These results confirm the conclusion made above of a direct interaction between the RBD of the EWS protein and the RNA-helicase p68 (Table 1).

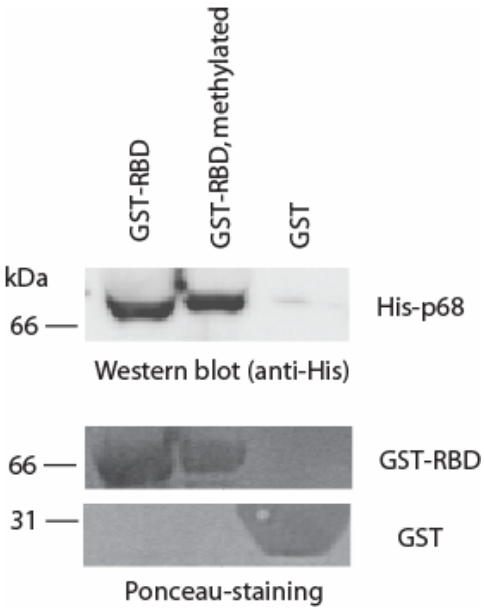


Figure 5.3: Purified His-p68 was used for a pull-down with GST-RBD as bait. Recombinant GST-RBD co-expressed with His-PRMT1 (methylated RBD) or not (unmethylated RBD), pre-bound to glutathione-coupled agarose, was incubated with purified His-p68. After extensive washing, eluted GST-RBD was subjected to SDS-PAGE. Bound recombinant His-p68 was detected by western blotting using an His-tag-specific antibody. Similar amounts of GST-RBD and even more in the GST control were used as shown by ponceau staining.

Co-localization of the EWS protein and RNA helicases p68 and p72

To investigate, whether the EWS protein and the RNA helicases p68 and p72 co-localize in cells, HEK 293 (T) cells were co-transfected and YFP-fused EWS protein and GST-p68 or GST-p72 were co-expressed. The EWS protein alone shows the expected nuclear localization with exclusion from the nucleoli (Fig. 5.4) (15). The RNA-helicases, detected by immunostaining using anti-GST antibodies, are also localized in the nucleus with minor fractions in the cytosol (Fig. 5.4). Endogenous GST of untransfected cells, which is also detected by the antibodies, was not present the nucleus (Fig. 5.4). When co-expressed, large parts of the EWS protein and the RNA helicases co-localize in the nucleus. Interestingly, upon co-expression of p68 or p72, the EWS protein changes its subnuclear distribution from nucleoplasm to the nucleolar periphery (Fig. 5.4).

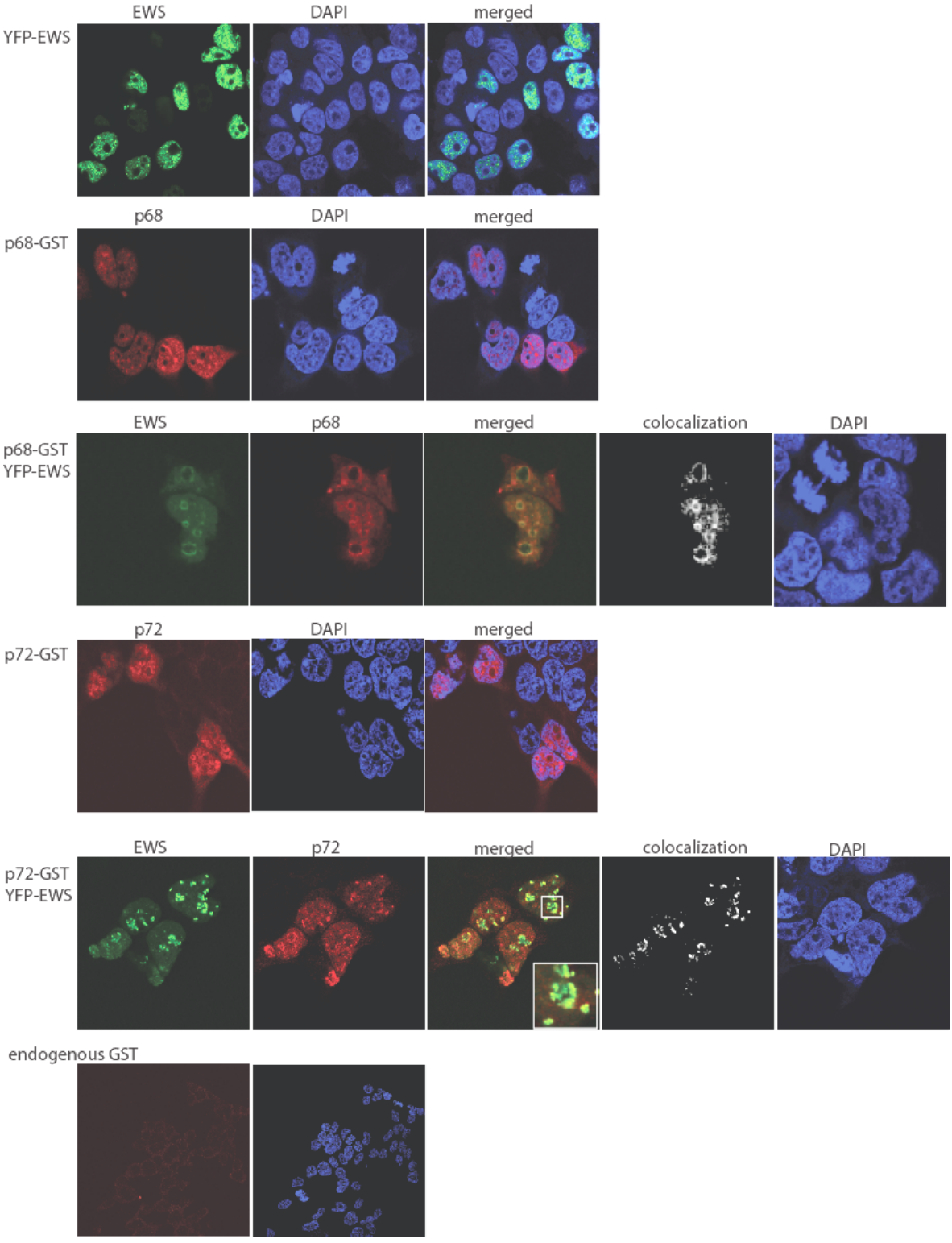


Figure 5.4: Co-localization of YFP-EWS and p68-GST or p72-GST. HEK cells were transiently transfected with YFP-EWS, p68-GST, or p72-GST alone, or co-transfected with YFP-EWS and p68, or YFP-EWS and p72-GST. P68-GST and p72-GST were visualized by immunofluorescence using a mouse anti-GST primary antibody and Indocarbocyanine (Cy3)-coupled goat anti-rabbit secondary antibodies. The fluorescent proteins were analyzed by confocal microscopy using excitation-emission wavelengths of 405-470 nm, 488-509 nm and 550-570 nm for DAPI, GFP and Cy3, respectively. The pictures were merged using the programs Adobe Photoshop and Imaris.

Methylation of the EWS protein occurs in the cytosol

By now, only methylated EWS protein could be found, be it at the cell surface or in the nucleus ((17) and own observations). The methyltransferase PRMT1, which methylates the EWS protein in vitro (21) was described to be localized either predominantly in the cytosol or in the nucleus (22-25). The observed discrepancy might depend on the different cell lines used and on different experimental setups, and reflect a dynamic subcellular distribution of PRMT1. Whether the EWS protein is methylated first in the cytoplasm and than transported to other cell compartments or the methylation takes place in the nucleus, and fractions of the methylated protein might be exported and transferred to the cell membrane, was still an open question.

To determine, where and when the methylation of the EWS protein occurs, HEK cells were transiently transfected with two different EWS constructs, His-EWS or a His-EWS-Y/F mutant, in which the last tyrosine of the EWS sequence was mutated to phenylalanine ($Y_{922}F$, corresponding to Y_{656} in the wild type EWS protein). This mutation in the nuclear localization signal (C-NLS) of the EWS protein prevents its nuclear import (15). The subcellular localization of the constructs were verified by fluorescence microscopy (Fig 5.5 A) confirming nuclear localization with exclusion from the nucleoli of His-EWS protein and accumulation of the His-EWS-Y/F mutant, designated here as cytoplasmic EWS protein, outside of the nucleus (Fig 5.5 B),. Thus, the methylation state of an EWS construct that was never in the nucleus can be compared with that present in the nucleus. The His-fusion proteins were purified using Ni-affinity agarose and subjected to SDS-PAGE. The EWS constructs, which were present as a dominant band at ~ 116 kDa on the coomassie-stained gel, were in-gel digested and analyzed by mass spectrometry. The peptide at m/z 1275.7 in the spectrum of the His-EWS construct covers the intact C-terminus (RQERRDRPY) and is absent in the spectrum of the His-EWS-Y/F mutant, whereas the peptide of the mutated C-terminus appeared at m/z 1259.7 (RQERRDRPF), confirming the mutation of the EWS C-terminus. In both cases, the tryptic and chymotryptic peptides cover large parts of the amino acid sequence of the constructs. From the 30 arginines, which were found to be methylated in the EWS protein in vivo and in vitro (17,21), 27 were asymmetrically dimethylated in the His-EWS wild-type construct, one arginine (R756) was found to be unmodified and the remaining two arginines (R569, R730) were not covered by the peptides. A similar methylation pattern was observed in the cytoplasmic His-EWS-Y/F mutant, 25 asymmetrically dimethylated arginines were detected, one arginine (R756) was found to be monomethylated, one arginine was

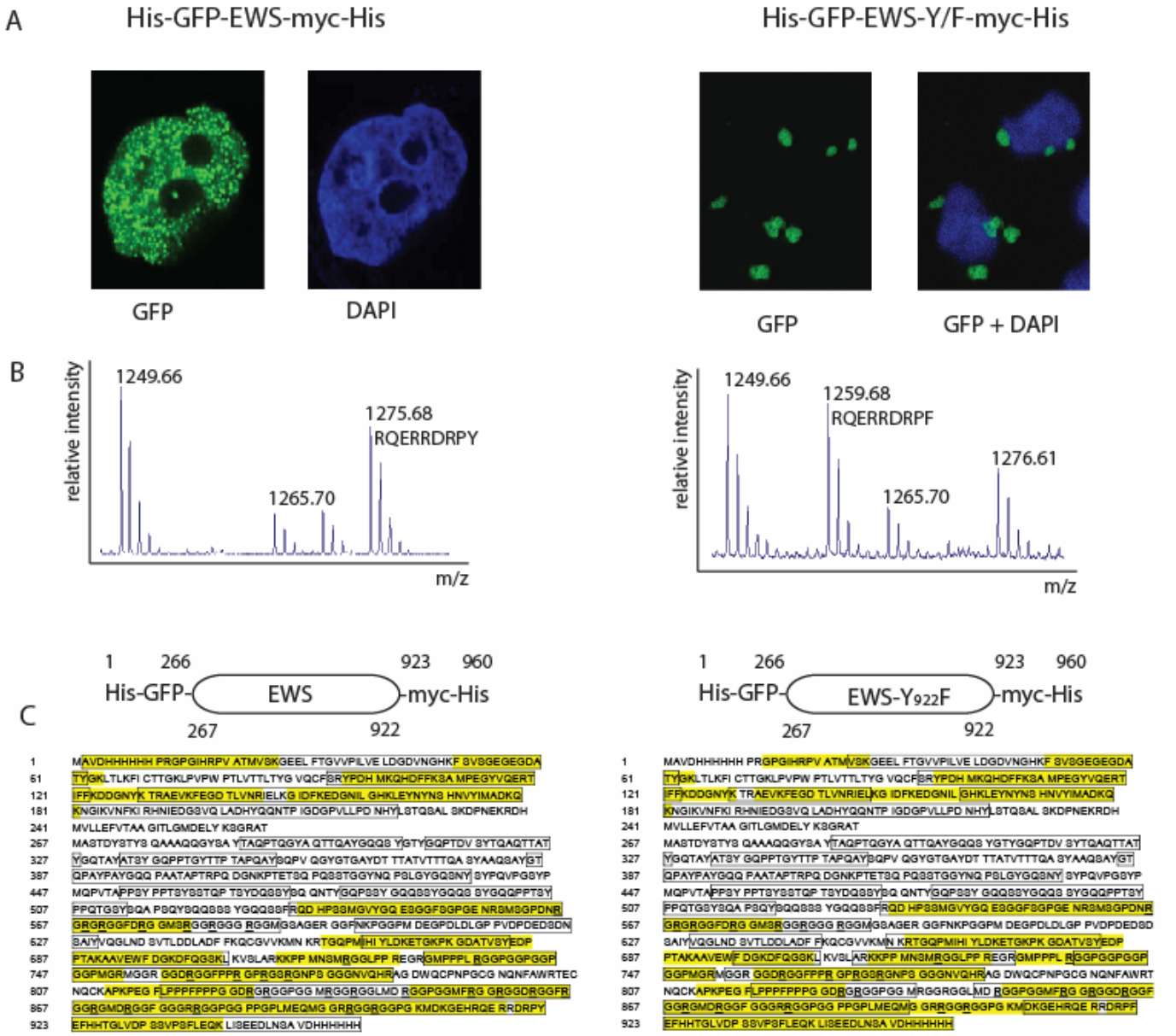


Figure 5.5: Methylation state of nuclear and cytoplasmic EWS protein. **(A)** HEK cells were transiently transfected with His-tagged GFP-EWS (nuclear) or GFP-EWS-Y/F (cytoplasmic) and fluorescent proteins were analyzed by confocal microscopy using excitation-emission wavelengths of 405-470 nm and 488-509 nm for DAPI and GFP, respectively. **(B)** The cells were lysed, the His-tagged EWS contructs were affinity-purified using Ni-affinity agarose, and extracted EWS protein was subjected to SDS-PAGE followed by in-gel digestion and MALDI-TOF-MS analysis. The spectrum shows the peptide covering the C-terminus of the EWS protein either without mutation (m/z 1275) or with the Y to F substitution (m/z 1259). **(C)** Peptide coverage of the His-EWS and the His-EWS-Y/F constructs analyzed by LC-MALDI-TOFTOF-MS. Large parts of the sequence is covered by the tryptic (box) or chymotryptic (yellow) peptides. Methylated (**R**) and dimethylated (**R**) arginines are labeled.

unmodified (R730), and the remaining three arginines (R569, R838, R841) were not covered by the peptides. Peptides in several methylation states were detected in both, nuclear and cytoplasmic His-EWS protein, which was not observed in the isolated wild type EWS protein

(own observations). The completely methylated peptides were always the dominant signal in the mass spectrum of the His-constructs. This incomplete methylation might be explained by the overexpression of the EWS constructs in the cells. All identified peptides including their modifications can be found in supplemental table 2.

Discussion

In this study, more than 30 proteins, which were pulled down by the EWS protein, could be unequivocally identified. Among them were mainly heterogeneous nuclear ribonucleoproteins (hnRNPs), but also RNA-helicases and structural proteins like actin, tubulin, and vimentin. Several methylated proteins were found among the identified proteins (Table 2) and for hnRNP H and $TAF_{II}68$, the identified methylation sites have not been, to our knowledge, described.

HnRNPs are highly abundant nuclear RNA-binding proteins which function in transcriptional regulation, mRNA splicing and processing, mRNA transport, and translation (26,27). Characteristic structural features of hnRNPs are RNA-recognition motifs (RRM), RGG-boxes containing asymmetrically dimethylated arginines, or KH-domains. In the nucleus, hnRNPs have been identified to form RNA-dependent complexes, most likely with (pre)-mRNA, consisting of ~20 major components with varying composition and stoichiometry. Although a function in splicing of several hnRNPs is proposed, they are no key elements of the spliceosome (28). The hnRNP-complex (H-complex) rather seems to bind free RNA or cover the intron sequences of the pre-mRNA to guide them to degradation in the nucleus (29). Here we showed that the EWS protein, methylated or unmethylated, pulled down an hnRNP-RNA complex. HnRNPs M and U have been demonstrated to be direct binding partners of the EWS protein, as their interaction is mediated even after digestion of RNA with RNase A. The C-terminal RBD of the EWS protein contains an RNA-recognition motif (RRM) and RGG-boxes with dimethylated arginines, structural features characterizing hnRNPs. Binding of the EWS RBD to the H-complex is independent of RBD methylation. As methylation of the hnRNPs is also not necessary for assembling the H-complex (30), and the EWS-RBD fulfills structural features of hnRNPs, interactions of the RBD, similar to hnRNPs, in the H-complex seem likely. However, transcriptional activation domains like the N-terminal EAD of the EWS protein are absent in hnRNPs. Therefore, the EWS protein might target hnRNP-complexes to their place of action in the cell. Considering the fact, that no core

proteins of the spliceosome were identified in this study, and that the EAD of the EWS protein is able to bind RNA polymerase II, it might function as scaffold protein in targeting hnRNPs and other RNA binding proteins to pre-mRNA already during transcription than playing a role in splicing.

Another direct interaction partner of the EWS protein, identified in this study, is the RNA-helicase p68 (DDX5). P68 and its homolog p72 (DDX17) are ATP-dependent RNA helicases located in the nucleus and, like the EWS protein, excluded from the nucleoli (31). They were identified as part of the human spliceosome in a large scale proteomic analysis (12) and a function as a promoter-specific transcriptional co-activator/repressor was reported (32,33). Recently, a role in ribosome biogenesis and cell proliferation was observed (34). In this study, we could show that the EWS protein interacts directly with p68 via the RBD. In HEK cells, we observed co-localization of the EWS protein with both p68 and p72. Interestingly, co-expression of the EWS protein and p68 or p72 changed the subnuclear distribution from nucleoplams, when expressed alone, to an accumulation in the periphery of the nuceoli, when co-expressed. Several nuceloplasmic RNA-binding proteins like PSF, p54, p68, TLS/FUS, and the EWS protein were identified to relocate to such nucleolar caps under transcriptional arrest (35), and a proteomic study of the nucleolar dynamics revealed an upregulation of p68, among others, in the nucleolus when transcription was inhibited (36). In that study, ~700 different proteins were identified as components of the nucleoli, but only one third are involved in rRNA biogenesis, an important function of the nucleolus. Therefore, an additional role in gene expression was proposed, as the nucleoli might serve as a place, where cellular regulators and their activity are sequestered (37) or even degraded, as proposed for the transcription factor c-myc (38).

Actin, tubulin, and vimentin were also found to interact with the EWS protein. These proteins are highly abundant in the cell and might be extracted unspecifically in such pull-down experiments, however, they were not found in any GST control pull-down indicating a functional relevance of these interactions. It was shown, that the EWS protein is present in RNA-transporting granules in dendrites, which transport their cargo via the microtubules (39) and in quiescent serum-starved cells, the EWS protein forms a net-like structure in the cytoplasm and is associated with microtubuli (Thesis of Rouzanna Zakaryan). Beside its cytoplasmic localization, actin was shown to be present as a monomer also in the nucleus (40). The role of the interaction with the EWS protein needs further investigations, but similarly to

the interaction of actin with hnRNP U in the nucleus, which is necessary for productive transcription of the RNA polymerase II (41), a functional relevance might be likely.

PRMT1, the main arginine methyltransferase in human cells, was also identified in the present study to bind to the unmethylated EWS protein in the presence and absence of RNase. As PRMT1 efficiently binds to and methylates the EWS protein in vitro and in vivo (21), its identification in this pull-down experiments confirms the specificity of the used method. Additionally, PRMT1 was found here to interact still with the methylated EWS protein. Even if all arginines are methylated, the methyltransferase PRMT1 seems to retain a high affinity towards the EWS protein, indicating a further role of PRMT1 as mediator than just being a methyltransferase. We could recently show that also PRMT8, a membrane-associated methyltransferase with high sequence similarity to PRMT1, interact with both, the methylated and unmethylated RGG3 box of the EWS protein (Pahlich et al., submitted for publication).

Binding partners of the EWS protein, which have been previously reported to interact with the EAD of the EWS protein, are the RNA polymerase II (1), calmodulin (42), the splicing factor 1 (43), or protein kinases (42,44). They were not found in this study, be it with the full length EWS protein or the EAD as bait. The EAD was predicted to be completely unstructured, a typical feature of human tumor-related transcription factors (45). Such intrinsically disordered proteins can mediate a stable secondary structure only upon ligand binding, as shown when CBP binds ACTR (46), or after post-translational modifications such as phosphorylation. Recombinant EWS protein or EAD might lack such ligands or modifications and therefore also a proper conformation for pulling down known and unknown interaction partners.

Additionally, it was demonstrated that the EWS protein is more or less completely methylated in the cytoplasm, most likely short after translation or even co-translationally, before it enters the nucleus. However, the binding of proteins and protein-RNA complexes to the EWS protein was not affected by this modification as was the transport of the EWS protein to the nucleus (15). Thus, it seems that the methylation of the EWS protein might be needed either for RNA binding, for activation/repression activity, or for stabilization, rather than being a regulatory element in its function as no unmethylated EWS protein was found in the cell and no demethylases are known so far.

References

1. Bertolotti, A., Melot, T., Acker, J., Vigneron, M., Delattre, O., and Tora, L. (1998) *Molecular and cellular biology* **18**(3), 1489-1497
2. Thomas, G. R., Faulkes, D. J., Gascoyne, D., and Latchman, D. S. (2004) *Biochem Biophys Res Commun* **318**(4), 1045-1051
3. Lee, J., Rhee, B. K., Bae, G. Y., Han, Y. M., and Kim, J. (2005) *Stem cells (Dayton, Ohio)* **23**(6), 738-751
4. Law, W. J., Cann, K. L., and Hicks, G. G. (2006) *Briefings in functional genomics & proteomics* **5**(1), 8-14
5. Delattre, O., Zucman, J., Plougastel, B., Desmaze, C., Melot, T., Peter, M., Kovar, H., Joubert, I., Dejong, P., Rouleau, G., Aurias, A., and Thomas, G. (1992) *Nature* **359**(6391), 162-165
6. Plougastel, B., Zucman, J., Peter, M., Thomas, G., and Delattre, O. (1993) *Genomics* **18**(3), 609-615
7. Alex, D., and Lee, K. A. (2005) *Nucleic Acids Res* **33**(4), 1323-1331
8. Yang, L., Embree, L. J., Tsai, S., and Hickstein, D. D. (1998) *J Biol Chem* **273**(43), 27761-27764
9. Yang, L., Chansky, H. A., and Hickstein, D. D. (2000) *J Biol Chem* **275**(48), 37612-37618
10. Chansky, H. A., Hu, M., Hickstein, D. D., and Yang, L. (2001) *Cancer Res* **61**(9), 3586-3590
11. Zinszner, H., Albalat, R., and Ron, D. (1994) *Genes Dev* **8**(21), 2513-2526
12. Rappsilber, J., Ryder, U., Lamond, A. I., and Mann, M. (2002) *Genome research* **12**(8), 1231-1245
13. Guipaud, O., Guillonneau, F., Labas, V., Praseuth, D., Rossier, J., Lopez, B., and Bertrand, P. (2006) *Proteomics* **6**(22), 5962-5972
14. Li, H., Watford, W., Li, C., Parmelee, A., Bryant, M. A., Deng, C., O'Shea, J., and Lee, S. B. (2007) *J Clin Invest*
15. Zakaryan, R. P., and Gehring, H. (2006) *Journal of molecular biology* **363**(1), 27-38
16. Lee, B. J., Cansizoglu, A. E., Suel, K. E., Louis, T. H., Zhang, Z., and Chook, Y. M. (2006) *Cell* **126**(3), 543-558
17. Belyanskaya, L. L., Gehrig, P. M., and Gehring, H. (2001) *Journal of Biological Chemistry* **276**(22), 18681-18687
18. Bedford, M. T., and Richard, S. (2005) *Molecular cell* **18**(3), 263-272
19. Pahlich, S., Zakaryan, R. P., and Gehring, H. (2006) *Biochimica et biophysica acta* **1764**(12), 1890-1903
20. Ong, S. E., Mittler, G., and Mann, M. (2004) *Nature methods* **1**(2), 119-126
21. Pahlich, S., Bschir, K., Chiavi, C., Belyanskaya, L., and Gehring, H. (2005) *Proteins* **61**(1), 164-175
22. Herrmann, F., Lee, J., Bedford, M. T., and Fackelmayer, F. O. (2005) *J Biol Chem* **280**(45), 38005-38010
23. Cote, J., Boisvert, F. M., Boulanger, M. C., Bedford, M. T., and Richard, S. (2003) *Mol Biol Cell* **14**(1), 274-287
24. Tang, J., Gary, J. D., Clarke, S., and Herschman, H. R. (1998) *Journal of Biological Chemistry* **273**(27), 16935-16945
25. Frankel, A., Yadav, N., Lee, J., Branscombe, T. L., Clarke, S., and Bedford, M. T. (2002) *J Biol Chem* **277**(5), 3537-3543
26. Krecic, A. M., and Swanson, M. S. (1999) *Current opinion in cell biology* **11**(3), 363-371
27. Dreyfuss, G., Kim, V. N., and Kataoka, N. (2002) *Nature reviews* **3**(3), 195-205
28. Bennett, M., Michaud, S., Kingston, J., and Reed, R. (1992) *Genes Dev* **6**(10), 1986-2000
29. Bennett, M., Pinol-Roma, S., Staknis, D., Dreyfuss, G., and Reed, R. (1992) *Molecular and cellular biology* **12**(7), 3165-3175
30. Pawlak, M. R., Banik-Maiti, S., Pietenpol, J. A., and Ruley, H. E. (2002) *J Cell Biochem* **87**(4), 394-407
31. Ogilvie, V. C., Wilson, B. J., Nicol, S. M., Morrice, N. A., Saunders, L. R., Barber, G. N., and Fuller-Pace, F. V. (2003) *Nucleic Acids Res* **31**(5), 1470-1480
32. Wilson, B. J., Bates, G. J., Nicol, S. M., Gregory, D. J., Perkins, N. D., and Fuller-Pace, F. V. (2004) *BMC molecular biology* **5**, 11
33. Bates, G. J., Nicol, S. M., Wilson, B. J., Jacobs, A. M., Bourdon, J. C., Wardrop, J., Gregory, D. J., Lane, D. P., Perkins, N. D., and Fuller-Pace, F. V. (2005) *Embo J* **24**(3), 543-553
34. Jalal, C., Uhlmann-Schiffler, H., and Stahl, H. (2007) *Nucleic Acids Res* **35**(11), 3590-3601
35. Shav-Tal, Y., Blechman, J., Darzacq, X., Montagna, C., Dye, B. T., Patton, J. G., Singer, R. H., and Zipori, D. (2005) *Mol Biol Cell* **16**(5), 2395-2413
36. Andersen, J. S., Lam, Y. W., Leung, A. K., Ong, S. E., Lyon, C. E., Lamond, A. I., and Mann, M. (2005) *Nature* **433**(7021), 77-83
37. Handwerger, K. E., and Gall, J. G. (2006) *Trends in cell biology* **16**(1), 19-26

38. Amati, B., and Sanchez-Arevalo Lobo, V. J. (2007) *Nature cell biology* **9**(7), 729-731
39. Kanai, Y., Dohmae, N., and Hirokawa, N. (2004) *Neuron* **43**(4), 513-525
40. Bettinger, B. T., Gilbert, D. M., and Amberg, D. C. (2004) *Nature reviews* **5**(5), 410-415
41. Kukalev, A., Nord, Y., Palmberg, C., Bergman, T., and Percipalle, P. (2005) *Nature structural & molecular biology* **12**(3), 238-244
42. Deloulme, J. C., Prichard, L., Delattre, O., and Storm, D. R. (1997) *J Biol Chem* **272**(43), 27369-27377
43. Zhang, D., Paley, A. J., and Childs, G. (1998) *J Biol Chem* **273**(29), 18086-18091
44. Felsch, J. S., Lane, W. S., and Peralta, E. G. (1999) *Curr Biol* **9**(9), 485-488
45. Iakoucheva, L. M., Brown, C. J., Lawson, J. D., Obradovic, Z., and Dunker, A. K. (2002) *Journal of molecular biology* **323**(3), 573-584
46. Demarest, S. J., Martinez-Yamout, M., Chung, J., Chen, H., Xu, W., Dyson, H. J., Evans, R. M., and Wright, P. E. (2002) *Nature* **415**(6871), 549-553

Supplemental Table 1: Proteins identified in GST pull-down experiments. Only proteins, which were not identified in any GST-control experiment, are listed. The accession numbers, the protein names, and the Mascot scores of the identifications are indicated.

Pull-down with GST-EWS as bait

prot_acc	prot_desc	Mascot scores Experiment 1	2	3	4	5
EWS_HUMAN	(Q01844) RNA-binding protein EWS	1158	505	703	555	10647
ANM1_HUMAN	(Q99873) Protein arginine N-methyltransferase 1			84	39	663
ROA2_HUMAN	(P22626) Heterogeneous nuclear ribonucleoproteins A2/B1	1625	644	566	663	
ROA1_HUMAN	(P09651) Heterogeneous nuclear ribonucleoprotein A1	1403	498	547	573	
HNRH1_HUMAN	(P31943) Heterogeneous nuclear ribonucleoprotein H	872	71	368		94
HNRH2_HUMAN	(P55795) Heterogeneous nuclear ribonucleoprotein H'	585		255	326	
ROA3_HUMAN	(P51991) Heterogeneous nuclear ribonucleoprotein A3	895	260	208	321	
HNRPU_HUMAN	(Q00839) Heterogeneous nuclear ribonucleoprotein U	65	185	164	49	257
HNRPG_HUMAN	(P38159) Heterogeneous nuclear ribonucleoprotein G	130		147	222	
HNRPM_HUMAN	(P52272) Heterogeneous nuclear ribonucleoprotein M			117	464	650
HNRPQ_HUMAN	(O60506) Heterogeneous nuclear ribonucleoprotein Q			93	73	
HNRPD_HUMAN	(Q14103) Heterogeneous nuclear ribonucleoprotein D0	50			110	
ROAA_HUMAN	(Q99729) Heterogeneous nuclear ribonucleoprotein A/B	32			55	
ROA0_HUMAN	(Q13151) Heterogeneous nuclear ribonucleoprotein A0	182		73	154	
HNRH3_HUMAN	(P31942) Heterogeneous nuclear ribonucleoprotein H3	442	98	52	268	
HNRPF_HUMAN	(P52597) Heterogeneous nuclear ribonucleoprotein F	723	67	21	249	
CIRBP_HUMAN	(Q14011) Cold-inducible RNA-binding protein (hnRNP A18)		64	48	134	
FUS_HUMAN	(P35637) RNA-binding protein FUS (hnRNP P)			216	233	
DDX5_HUMAN	(P17844) Probable RNA-dependent helicase p68	230		255	252	
DDX17_HUMAN	(Q92841) Probable RNA-dependent helicase p72	104		171	142	
DDX3X_HUMAN	(O00571) DEAD-box protein 3, X-chromosomal	101		118	112	
DHX9_HUMAN	(Q08211) ATP-dependent RNA helicase A			105	55	82
Q8IYE5	(Q8IYE5) ATP-dependent RNA helicase DHX36			67	35	
SFRS3_HUMAN	(P84103) Splicing factor, arginine/serine-rich 3	60		80	64	
SFRS9_HUMAN	(Q13242) Splicing factor, arginine/serine-rich 9	117	54		133	
SFRS1_HUMAN	(Q07955) Splicing factor, arginine/serine-rich 1	280	40		123	
SMD3_HUMAN	(P62318) Small nuclear ribonucleoprotein Sm D3	58			75	
SNRPA_HUMAN	(P09012) U1 small nuclear ribonucleoprotein A					
RBP56_HUMAN	(Q92804) TATA-binding protein-associated factor 2N (TAFII68)	63		131		
RS18_HUMAN	(P62269) 40S ribosomal protein S18				31	31
EF1G_HUMAN	(P26641) Elongation factor 1-gamma			33		234
EF1A1_HUMAN	(P68104) Elongation factor 1-alpha 1				88	373
Q96PK6	(Q96PK6) Coactivator activator			67	67	
ACTB_HUMAN	(P60709) Actin, cytoplasmic 1 (Beta-actin)	218	213	78	174	1172
TBAK_HUMAN	(P68363) Tubulin alpha-ubiquitous chain	39	53	51	61	691
TBB2_HUMAN	(P07437) Tubulin beta-2 chain	62		48	83	1236
VIME_HUMAN	(P08670) Vimentin		43			98
RS19_HUMAN	(P39019) 40S ribosomal protein S19					29
RL27A_HUMAN	(P46776) 60S ribosomal protein L27a					23
RS5_HUMAN	(P46782) 40S ribosomal protein S5					83
RS10_HUMAN	(P46783) 40S ribosomal protein S10					123
RS7_HUMAN	(P62081) 40S ribosomal protein S7					26
RS15A_HUMAN	(P62244) 40S ribosomal protein S15a					209
RS16_HUMAN	(P62249) 40S ribosomal protein S16					202
RS14_HUMAN	(P62263) 40S ribosomal protein S14					31
RS18_HUMAN	(P62269) 40S ribosomal protein S18 (Ke-3) (Ke3)					481
RS11_HUMAN	(P62280) 40S ribosomal protein S11					51
RL23_HUMAN	(P62829) 60S ribosomal protein L23 (Ribosomal protein L17)					29
RS25_HUMAN	(P62851) 40S ribosomal protein S25					53
RL38_HUMAN	(P63173) 60S ribosomal protein L38					43
Q5VVD0	(Q5VVD0) Ribosomal protein L11					136
RLA1_HUMAN	(P05386) 60S acidic ribosomal protein P1					75
RLA0_HUMAN	(P05388) 60S acidic ribosomal protein P0 (L10E)					74
RS3_HUMAN	(P23396) 40S ribosomal protein S3					303
RS4X_HUMAN	(P62701) 40S ribosomal protein S4, X isoform					50
Q3MIH3	(Q3MIH3) Ubiquitin and ribosomal protein L40,					29
Q5VXM2	(Q5VXM2) Mitochondrial ribosomal protein L37					97

Supplemental Table 1: (continued)

Pull-down with GST-EWS as bait + Rnase A

prot_acc	prot_desc	Mascot scores Experiment 1	2	3
EWS_HUMAN	(Q01844) RNA-binding protein EWS	1158	471	12334
HNRPM_HUMAN	(P52272) Heterogeneous nuclear ribonucleoprotein M		36	466
FUS_HUMAN	(P35637) RNA-binding protein FUS (hnRNP P)	121	84	
DDX5_HUMAN	(P17844) Probable RNA-dependent helicase p68	143	217	54
DDX17_HUMAN	(Q92841) Probable RNA-dependent helicase p72	166	211	
DDX3X_HUMAN	(O00571) DEAD-box protein 3, X-chromosomal	148	241	
TBB2_HUMAN	(P07437) Tubulin beta-2 chain	392		1752
ACTB_HUMAN	(P60709) Actin, cytoplasmic 1 (Beta-actin)	247		1339
RS3_HUMAN	(P23396) 40S ribosomal protein S3			289
Q5VXM2	(Q5VXM2) Mitochondrial ribosomal protein L37 (Fragment)			193
RS13_HUMAN	(P62277) 40S ribosomal protein S13			180
RLA0_HUMAN	(P05388) 60S acidic ribosomal protein P0 (L10E)			174
RS4X_HUMAN	(P62701) 40S ribosomal protein S4			166
RS3A_HUMAN	(P61247) 40S ribosomal protein S3a			158
RL21_HUMAN	(P46778) 60S ribosomal protein L21			137
RL13_HUMAN	(P26373) 60S ribosomal protein L13			130
Q6IPX4	(Q6IPX4) RPS16 protein			107
RS9_HUMAN	(P46781) 40S ribosomal protein S9			105
Q5VVC8	(Q5VVC8) Ribosomal protein L11 (Fragment)			97
RS7_HUMAN	(P62081) 40S ribosomal protein S7			85
RS15_HUMAN	(P62841) 40S ribosomal protein S15 (RIG protein)			77
Q5JR95	(Q5JR95) Ribosomal protein S8			76
RL4_HUMAN	(P36578) 60S ribosomal protein L4 (L1)			75
RS18_HUMAN	(P62269) 40S ribosomal protein S18 (Ke-3) (Ke3)			72
RL3_HUMAN	(P39023) 60S ribosomal protein L3			71
Q6IPH7	(Q6IPH7) RPL14 protein (Ribosomal protein L14 variant)			67
RL27A_HUMAN	(P46776) 60S ribosomal protein L27a			57
Q4VXZ3	(Q4VXZ3) Ribosomal protein S18			45
RL18A_HUMAN	(Q02543) 60S ribosomal protein L18a			40
Q3MIH3	(Q3MIH3) Ubiquitin and ribosomal protein L40			30
RS11_HUMAN	(P62280) 40S ribosomal protein S11			30
RL5_HUMAN	(P46777) 60S ribosomal protein L5			30

Supplemental Table 1: (continued)

Pull-down with methylated GST-EWS as bait

prot_acc	prot_desc	Mascot scores Experiment 1	2	3
EWS_HUMAN	(Q01844) RNA-binding protein EWS	873	3463	8595
ANM1_HUMAN	(Q99873) Protein arginine N-methyltransferase 1	98	507	161
FUS_HUMAN	(P35637) RNA-binding protein FUS (hnRNP P)	69	127	35
HNRH1_HUMAN	(P31943) Heterogeneous nuclear ribonucleoprotein H	76	218	40
HNRPD_HUMAN	(Q14103) Heterogeneous nuclear ribonucleoprotein D0	39	73	
HNRPG_HUMAN	(P38159) Heterogeneous nuclear ribonucleoprotein G	39	193	
HNRPM_HUMAN	(P52272) Heterogeneous nuclear ribonucleoprotein M	187		976
HNRPU_HUMAN	(Q00839) Heterogenous nuclear ribonucleoprotein U	260	118	23
Q5T0N2	(Q5T0N2) Heterogeneous nuclear ribonucleoprotein F	63	252	22
ROA1_HUMAN	(P09651) Heterogeneous nuclear ribonucleoprotein A1	182	786	
ROA2_HUMAN	(P22626) Heterogeneous nuclear ribonucleoproteins A2/B1	179	441	
ROA3_HUMAN	(P51991) Heterogeneous nuclear ribonucleoprotein A3	130	300	
DDX5_HUMAN	(P17844) Probable RNA-dependent helicase p68	160	526	61
DDX3X_HUMAN	(O00571) DEAD-box protein 3, X-chromosomal	65	452	
DHX9_HUMAN	(Q08211) ATP-dependent RNA helicase A	17	135	
CPSF5_HUMAN	(O43809) Cleavage and polyadenylation specificity factor 5	35	41	
EF1A1_HUMAN	(P68104) Elongation factor 1-alpha 1	63	143	412
EF1G_HUMAN	(P26641) Elongation factor 1-gamma	50	43	38
EFTU_HUMAN	(P49411) Elongation factor Tu		127	99
H2A1B_HUMAN	(P04908) Histone H2A type 1-B	63		58
H4_HUMAN	(P62805) Histone H4	22	302	178
MCM7_HUMAN	(P33993) DNA replication licensing factor MCM7		33	165
SMD3_HUMAN	(P62318) Small nuclear ribonucleoprotein Sm D3	37	54	42
ACTB_HUMAN	(P60709) Actin, cytoplasmic 1 (Beta-actin)	115	554	777
TBAK_HUMAN	(P68363) Tubulin alpha-ubiquitous chain	109	229	335
TBB2_HUMAN	(P07437) Tubulin beta-2 chain	191	577	834
TBB2C_HUMAN	(P68371) Tubulin beta-2C chain	188	415	830
VIME_HUMAN	(P08670) Vimentin		186	161
ADT2_HUMAN	(P05141) ADP/ATP translocase 2	59		122
ATPA_HUMAN	(P25705) ATP synthase alpha chain	87	302	478
ATPB_HUMAN	(P06576) ATP synthase beta chain		112	37
ATPO_HUMAN	(P48047) ATP synthase O subunit	45		95
ENOA_HUMAN	(P06733) Alpha-enolase		253	37
G3P_HUMAN	(P04406) Glyceraldehyde-3-phosphate dehydrogenase	100	188	108
GRP75_HUMAN	(P38646) Stress-70 protein, mitochondrial precursor		64	238
HSP7C_HUMAN	(P11142) Heat shock cognate 71 kDa protein	68	63	628
IDHP_HUMAN	(P48735) Isocitrate dehydrogenase [NADP]		115	126
LDHB_HUMAN	(P07195) L-lactate dehydrogenase B chain		197	69
C1TC_HUMAN	(P11586) C-1-tetrahydrofolate synthase, cytoplasmic		59	97
IREB1_HUMAN	(P21399) Iron-responsive element-binding protein 1		68	26
RAC2_HUMAN	(P15153) Ras-related C3 botulinum toxin substrate 2 precursor		54	54
SERA_HUMAN	(O43175) D-3-phosphoglycerate dehydrogenase		83	19
Q5JR95	(Q5JR95) Ribosomal protein S8		177	75
Q5T4L4	(Q5T4L4) Ribosomal protein S27		73	99
Q5VVD0	(Q5VVD0) Ribosomal protein L11	71	86	154
Q5VZM5	(Q5VZM5) Ribosomal protein L21	124	142	
RL10A_HUMAN	(P62906) 60S ribosomal protein L10a (CSA-19)		94	53
RL13_HUMAN	(P26373) 60S ribosomal protein L13		179	77
RL17_HUMAN	(P18621) 60S ribosomal protein L17 (L23)	21	117	22
RL18A_HUMAN	(Q02543) 60S ribosomal protein L18a		62	47
RL27A_HUMAN	(P46776) 60S ribosomal protein L27a		69	70
RL3_HUMAN	(P39023) 60S ribosomal protein L3		42	86
RL31_HUMAN	(P62899) 60S ribosomal protein L31	25		34
RLA0_HUMAN	(P05388) 60S acidic ribosomal protein P0 (L10E)	103		115
RS10_HUMAN	(P46783) 40S ribosomal protein S10		106	156
RS11_HUMAN	(P62280) 40S ribosomal protein S11	23		127
RS13_HUMAN	(P62277) 40S ribosomal protein S13		118	82
RS15A_HUMAN	(P62244) 40S ribosomal protein S15a	57	130	212
RS16_HUMAN	(P62249) 40S ribosomal protein S16		136	482
RS17_HUMAN	(P08708) 40S ribosomal protein S17	86		40
RS18_HUMAN	(P62269) 40S ribosomal protein S18 (Ke-3) (Ke3)	27	163	835
RS19_HUMAN	(P39019) 40S ribosomal protein S19	21	86	48
RS2_HUMAN	(P15880) 40S ribosomal protein S2 (S4) (LLRep3 protein)		38	37
RS3_HUMAN	(P23396) 40S ribosomal protein S3	154	179	309
RS4X_HUMAN	(P62701) 40S ribosomal protein S4, X isoform		74	141
RS7_HUMAN	(P62081) 40S ribosomal protein S7	41	78	72

Supplemental Table 1: (continued)

Pull-down with methylated GST-EWS as bait + Rnase A

prot_acc	prot_desc	Mascot scores Experiment 1	2	3
EWS_HUMAN	(Q01844) RNA-binding protein EWS	832	1472	9801
ANM1_HUMAN	(Q99873) Protein arginine N-methyltransferase 1		666	603
SKB1_HUMAN	(O14744) Protein arginine N-methyltransferase 5		73	120
HNRPM_HUMAN	(P52272) Heterogeneous nuclear ribonucleoprotein M	34		788
HNRPU_HUMAN	(Q00839) Heterogenous nuclear ribonucleoprotein U		125	437
DDX5_HUMAN	(P17844) Probable RNA-dependent helicase p68		200	45
EF1A2_HUMAN	(Q05639) Elongation factor 1-alpha 2		54	248
EF1B_HUMAN	(P24534) Elongation factor 1-beta		101	191
EF1G_HUMAN	(P26641) Elongation factor 1-gamma	44	46	209
MCM7_HUMAN	(P33993) DNA replication licensing factor MCM7		33	126
SMD3_HUMAN	(P62318) Small nuclear ribonucleoprotein Sm D3		91	19
ACTB_HUMAN	(P60709) Actin, cytoplasmic 1 (Beta-actin)	143	476	1109
TBAK_HUMAN	(P68363) Tubulin alpha-ubiquitous chain	219	282	603
TBB2_HUMAN	(P07437) Tubulin beta-2 chain	59	696	753
TBB2C_HUMAN	(P68371) Tubulin beta-2C chain	61	659	820
TCPA_HUMAN	(P17987) T-complex protein 1 subunit alpha		141	45
ACPH_HUMAN	(P13798) Acylamino-acid-releasing enzyme		60	22
ATPA_HUMAN	(P25705) ATP synthase alpha chain		114	236
C1TC_HUMAN	(P11586) C-1-tetrahydrofolate synthase, cytoplasmic		44	251
DCD_HUMAN	(P81605) Dermcidin precursor		32	73
DIRA2_HUMAN	(Q96HU8) GTP-binding protein Di-Ras2		21	35
G3P_HUMAN	(P04406) Glyceraldehyde-3-phosphate dehydrogenase	40	241	110
HSP7C_HUMAN	(P11142) Heat shock cognate 71 kDa protein	108	119	936
IDHP_HUMAN	(P48735) Isocitrate dehydrogenase [NADP]		93	83
LDHB_HUMAN	(P07195) L-lactate dehydrogenase B chain		192	68
PSD3_HUMAN	(O43242) 26S proteasome non-ATPase regulatory subunit 3		25	120
Q5CAQ7	(Q5CAQ7) Heat shock protein HSP 90-alpha 2		277	246
Q6NZ56	(Q6NZ56) Hypothetical protein		40	74
Q9NTK6	(Q9NTK6) Hypothetical protein DKFZp761K0511		429	244
RIB1_HUMAN	(P04843) Dolichyl-diphosphooligosaccharide--protein glycosyltransferase		25	83
SERA_HUMAN	(O43175) D-3-phosphoglycerate dehydrogenase		83	79
TFR1_HUMAN	(P02786) Transferrin receptor protein 1		58	48
Q5VVD0	(Q5VVD0) Ribosomal protein L11		47	127
RL13_HUMAN	(P26373) 60S ribosomal protein L13		104	83
RL18A_HUMAN	(Q02543) 60S ribosomal protein L18a		26	51
RL24_HUMAN	(P83731) 60S ribosomal protein L24		80	74
RLA0_HUMAN	(P05388) 60S acidic ribosomal protein P0 (L10E)		123	80
RS13_HUMAN	(P62277) 40S ribosomal protein S13		74	149
RS15A_HUMAN	(P62244) 40S ribosomal protein S15a		113	221
RS16_HUMAN	(P62249) 40S ribosomal protein S16		201	170
RS18_HUMAN	(P62269) 40S ribosomal protein S18 (Ke-3) (Ke3)		126	223
RS3_HUMAN	(P23396) 40S ribosomal protein S3		34	293
RS3A_HUMAN	(P61247) 40S ribosomal protein S3a		43	307
RS4X_HUMAN	(P62701) 40S ribosomal protein S4, X isoform		112	225
RS9_HUMAN	(P46781) 40S ribosomal protein S9		88	78

Acknowledgements

I would like to thank Heinz Gehring for giving me the opportunity to work in his lab on the interesting project about the EWS protein. I am grateful for his continuous support, the scientific freedom I enjoyed, and his willingness to be available for questions and discussions at any time.

I further want to thank Philipp Christen for his interest in my project, his expert advices in experimental questions, as well as his critical and helpful comments and discussions during the seminars and manuscript preparations.

I am grateful to Prof. Dr. Sonderegger for supervising my thesis as responsible faculty member.

I am thankful to Peter Hunziker and Serge Chesnov from the Protein Analysis Unit for allowing me to use the MALDI-TOF-MS independently whenever it was free, and for their open-minded help and support in any questions regarding mass spectrometry. Peter Gehrig, Bernd Roschitzki, Bertran Gerrits, and Mike Scott from the Functional Genomics Centre (FGCZ) I want to thank for their support in recording and analyzing my data within the FGCZ facility.

I am grateful to my colleagues in the lab, Doris Grossenbacher, Rouzanna Zakaryan, Fang Wu, and the diploma and bachelor students Maria Jose Sedda, Lilan Quero, and Iwan Zimmermann for their spontaneous help, the friendly and collaborative working atmosphere, the funny time in the lab, and the enjoyable breaks outside the lab, for which I also want to thank Rahel, Christine, Anja, Marco, Wojti, Min-Li, and Po-Chi.

Finally, I want to thank my family and especially Gudi Speidel for the support, patience, and appreciation within the last years.

CURRICULUM VITAE

PAHLICH, Steffen

Diplom-Biotechnologe

Geboren am	29.03.1977 in Giessen, Deutschland
Staatsangehörigkeit:	deutsch

Ausbildung:

1983 - 1987	Grundschule, Lich, Deutschland
1987 - 1993	Gesamtschule, Lich, Deutschland
1993 - 1996	Gymnasiale Oberstufe, Grünberg, Deutschland
Juli 1996	Erlangung der Allgemeinen Hochschulreife (Abitur)
Okt. 1997	Studium der Biologie an der TU Braunschweig, Deutschland
April 1998	Wechsel zum Studium der Biotechnologie an der TU Braunschweig, Deutschland
Aug. 1999	Vordiplom in Biotechnologie
Aug. 2002 – Jan. 2003	Externe Diplomarbeit bei der Firma Roche Vitamins, Abteilung Biotechnologie, in Kaiseraugst, Schweiz. “Characterisierung von Post-Translationalen Modifikationen in *Bacillus subtilis* mit Methoden der Proteomanalyse”.
Feb. 2003	Diplom in Biotechnologie (Dipl. Biotechnol.) an der Technischen Universität (TU) Braunschweig, Deutschland
2003 - 2007	Assistent am Biochemischen Institut der Universität Zürich, Schweiz

An der Universität Zürich seit September 2003

Titel der Dissertation:	“Functional Characterization of the Ewing Sarcoma (EWS) Protein and its Post-Translational Modifications”
Leiter der Dissertation:	Prof. Dr. Heinz Gehring
Gegenwärtige Stellung:	Assistent / Doktorand